

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS DE LA INFORMACIÓN



**PREDICCIÓN ELECTORAL MEDIANTE ANÁLISIS
DE REDES SOCIALES**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR**

Montserrat Fernández Crespo

Bajo la dirección del doctor

Antonio López Vega

MADRID, 2013

Predicción Electoral

Mediante Análisis de Redes Sociales

Autora: Montserrat Fernández Crespo

Director: Antonio López Vega

Programa: Doctorado en Problemas Contemporáneos en la Sociedad de la Información
(IUIOG-Universidad Complutense)

Noviembre de 2012

A Alex, mi razón.

Agradecimientos

A Mariluz Congosto (@conghosto), doctoranda de la Universidad Carlos III de Madrid, experta en minería de datos Twitter, la persona que me ha facilitado los datos para la investigación. Sin su generosidad, esta Tesis Doctoral no podría haberse realizado.

A Alberto Mora (@AlbertoMora), politólogo y compañero, por haberme ayudado enormemente en la definición de la metodología del análisis cuantitativo. Sus sabios consejos en el tratamiento de datos han resultado imprescindibles.

A Daniel Gayo-Avello (@PFCdgayo), profesor de la Universidad de Oviedo (España), y a Pedro Magalhães (@PCMagalhaes), del Social Sciences Institute of the University of Lisbon (ICS), (Portugal). A ambos porque sin conocerme, respondieron a mis correos remitiéndome amablemente sus *papers*.

A mi tutor, Antonio López Vega (@antoniolvega), por su confianza en mi capacidad investigadora, sus ánimos constantes y su infinita paciencia en la corrección de este trabajo.

Al Instituto Universitario de Investigación Ortega y Gasset, y en especial a Ismael Crespo, Director del programa de doctorado, por procurarme un entorno investigador.

A Internet, y a todas las personas que en ella participan.

Índice

1.	Introducción	13
2.	Capítulo I: El comportamiento de las Redes Sociales	35
2.1.	Introducción	35
2.2.	Conociendo las Redes Sociales	36
2.2.1.	El concepto de Red Social	36
2.2.2.	Propiedades de las redes sociales	38
2.2.3.	Las reglas de la red	42
2.3.	Teorías de redes sociales	43
2.3.1.	La importancia de la centralidad	43
2.3.2.	El mundo pequeño de Milgram (Seis grados de separación)	45
2.3.3.	Tres grados de influencia	48
2.3.4.	La fuerza de los vínculos débiles	49
2.3.5.	El contagio en la red: memes y difusión cultural	52
2.4.	El comportamiento de Internet como red	60
2.4.1.	Características de Internet según teorías de redes	60
2.4.2.	Facebook y Twitter, nuevas redes	67
2.4.3.	El contagio en Facebook	69
2.4.4.	El contagio en Twitter	82
2.5.	Política y red	95
2.5.1.	El contagio político	95
2.5.2.	Internet y la polarización de grupos	103
2.6.	Conclusiones	112
3.	Capítulo II: Marco Teórico	123

3.1.	Introducción	123
3.2.	El auge de la política en Internet.....	125
3.3.	Antecedentes en el uso del análisis de Redes Sociales como herramienta de predicción.....	131
3.3.1.	Facebook como herramienta de predicción electoral	138
3.3.2.	Google Trends como herramienta de predicción electoral	148
3.3.3.	Twitter como herramienta de predicción electoral	153
3.3.4.	Análisis de mercados de apuestas como predicción electoral	173
3.3.5.	El peso de las Ciber campañas en el éxito electoral	179
3.4.	Encuestas preelectorales. Eficacia y desviación	185
3.5.	Penetración de las Redes Sociales en la población española.....	197
3.6.	Conclusiones	209
4.	Capítulo III: Metodología.....	217
4.1.	Introducción	217
4.2.	Metodología para el análisis	221
4.2.1.	Primera línea de análisis: toma de datos en la campaña electoral	221
4.2.2.	Segunda línea de análisis: estimación de la ponderación de las variables	225
4.2.3.	Propuesta de tercera línea de análisis: diferentes periodos en la toma de datos	230
4.3.	Metodología para la captura de datos	232
4.3.1.	Elecciones Generales 2011	233
4.3.2.	Elecciones Autonómicas a la Comunidad de Madrid 2011	234
4.3.3.	Elecciones Autonómicas a la Región de Murcia 2011	234
4.3.4.	Elecciones al Parlamento de Cataluña 2010.....	235
4.4.	Acerca de la validez de la muestra.....	236

5. Capítulo IV: Resultados de la investigación	243
5.1. Introducción	243
5.2. Análisis primera y segunda línea de investigación	246
5.2.1. Elecciones Generales 2011	246
5.2.1.1. Datos de Partida Elecciones Generales 2011	246
5.2.1.2. Resultados de la Investigación Elecciones Generales 2011	247
5.2.1.3. Conclusiones Análisis Twitter en las Elecciones Generales 2011	251
5.2.1.4. Encuestas Electorales versus Análisis Twitter en las Elecciones Generales 2011	252
5.2.2. Elecciones Autonómicas a la Comunidad de Madrid 2011	254
5.2.2.1. Datos de Partida Elecciones Autonómicas a la Comunidad de Madrid 2011	254
5.2.2.2. Resultados de la Investigación Elecciones Autonómicas a la Comunidad de Madrid 2011	256
5.2.2.3. Conclusiones Análisis Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011	259
5.2.2.4. Encuestas Electorales versus Análisis Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011	260
5.2.3. Elecciones Autonómicas a la Región de Murcia 2011	262
5.2.3.1. Datos de Partida Elecciones Autonómicas a la Región de Murcia 2011	262
5.2.3.2. Resultados de la Investigación Elecciones Autonómicas a la Región de Murcia 2011	264
5.2.3.3. Conclusiones Análisis Twitter en las Elecciones Autonómicas a la Región de Murcia 2011	268
5.2.3.4. Encuestas Electorales versus Análisis Twitter en las Elecciones Autonómicas a la Región de Murcia 2011	269
5.2.4. Elecciones al Parlamento de Cataluña 2010	272
5.2.4.1. Datos de Partida Elecciones al Parlamento de Cataluña 2010	272
5.2.4.2. Resultados de la Investigación Elecciones al Parlamento de Cataluña 2010	274

5.2.4.3. Conclusiones Análisis Twitter en las Elecciones al Parlamento de Cataluña 2010	275
5.2.4.4. Encuestas Electorales versus Análisis Twitter en las Elecciones al Parlamento de Cataluña 2010.....	276
5.3. Una Aproximación Metodológica: Resultado del Análisis según Tercera Línea de Investigación.....	279
5.4. Twitter y el resultado electoral: posibles causas de eficacia y desviación	282
5.5. Conclusiones.....	286
6. Conclusiones	295
7. Bibliografía.....	311
7.1. Obras.....	311
7.2. Papers.....	312
7.3. Artículos y consultas Web	319
7.4. Informes	322
8. Índice de ilustraciones y tablas	327
8.1. Índice de ilustraciones	327
8.2. Índice de tablas	330

Introducción

1. Introducción

El análisis de la teoría de redes sociales, desempeñado por antropólogos, politólogos y sociólogos, entre otros, ha intentado comprender el funcionamiento de una red social, buscando patrones de comportamiento en sus actores –nodos- y en las relaciones -vínculos- que se establecen entre ellos. Son muchos los estudios dedicados al diseño de su morfología y a la definición de sus características y propiedades, tratando de identificar los rasgos fundamentales que influyen en el papel que juega cada actor en la red y los patrones de difusión y contagio de todo aquello que circula por la misma.

En política también se da el contagio. El análisis racional de ir a votar resultó insuficiente porque la teoría de la racionalidad pasaba por alto una particularidad esencial: “no votamos solos”, el voto de un individuo puede propagarse como el fuego desencadenando una cascada. Así, nuestra decisión de votar influye, y siguiendo la regla de los Tres Grados de Influencia, podría extenderse desde nuestros amigos, hasta los amigos de éstos, y los amigos de los amigos de nuestros amigos. Consideramos que si aceptamos que la intención de voto es un acto de naturaleza contagiosa y las redes sociales en Internet son capaces de manifestar características propias –la influencia de nodos no centrales, la longitud y forma de la cadena de difusión, multiplicación de conexiones, proliferación de vínculos débiles- que difieren de las teorizadas en las redes tradicionales, en este canal, el contagio político puede adquirir valores numéricos y comportamientos que sorprenderían por su cantidad y procedimientos únicos.

¿Por qué votamos si nuestro voto cuenta poco? ¿Qué nos mueve a hacerlo cuándo intuimos que nuestra participación, por sí misma, no es decisiva? ¿Decidimos solos, nos dejamos influir o pretendemos influenciar a los demás? ¿Qué nos empuja a manifestar nuestra opinión política en Internet? La razón de que votemos, en realidad, tiene mucho que ver con nuestra pertenencia a grupos de sociedad y con el poder de las redes sociales, físicas o virtuales. (Christakis y Fowler, 2010)

Además de revisar algunas teorías sobre el mecanismo de contagio político entre individuos, nos preguntamos por la capacidad de éxito en el uso de redes sociales en una campaña electoral y cuáles son las líneas de actuación que marcan la diferencia. ¿Cómo decidió la gente sumarse a la campaña de Barack Obama en Estados Unidos en 2008 cuando todavía no se había extendido la percepción del candidato como posibilidad? ¿Cómo logró persuadir a tanta gente que hasta entonces no se había implicado en política? Obama triunfó en gran medida porque muchas personas se sintieron conectadas. Se ha comentado la extraordinaria capacidad de Obama de conectar con los votantes, pero más impresionante aún fue su capacidad de conectar a unos votantes con otros.

El análisis de campañas electorales online ha manifestado que la red tiene capacidad de influir en un efecto positivo en los resultados electorales y que los candidatos con mejor posicionamiento en las redes sociales pueden aventajar a sus oponentes. Pero la realidad no es tan básica ni fácil de prever en este campo. Algunos candidatos preferidos en redes sociales como Facebook o Twitter han sido derrotados por sus oponentes, consiguiendo, no obstante, mejorar su posicionamiento respecto a anteriores contiendas. Para llegar a estudios de los que se desprendan conclusiones que ayuden a marcar una posible estrategia electoral en Internet, sería necesario analizar la gran cantidad de detalles que se orquestan en la Red alrededor de una campaña electoral y de sus candidatos, referidos en su mayoría a cómo se difunden los mensajes emitidos por el candidato y la cualidad, positiva, negativa o neutra, de las reacciones que provoca: se comparten los mensajes, hasta dónde llegan en la cadena, quién es el actor idóneo para re-difundirlos, cuántas acciones particulares fuera de su propias páginas consigue, cuáles son los mensajes que más atraen, y cuántas y cómo son las reacciones que consigue cada uno de ellos, entre otras.

Desde la campaña de Barack Obama en 2008, en la que su apuesta por Internet ha sido definida como el punto clave de su victoria electoral, derrotando primero a Hillary Clinton en las primarias y posteriormente a McCain en las

generales, ha quedado más que patente la convicción de que cualquier candidato que quiera conseguir una mejoría en sus resultados e incluso un éxito electoral, está prácticamente obligado a desarrollar una cibercampaña.

Si bien es cierto que Obama destinó una cantidad mayor de recursos en Internet que sus oponentes, también lo es que gastó 10 veces más, según los datos oficiales de la Federal Election Commission, que Clinton en publicidad digital en Google, Yahoo y Facebook. Además, otros factores contribuyeron a su victoria: la capacidad de organizar a ciberactivistas que trasladan su actividad al “mundo real” y el hecho de el 74% de los seguidores online de Obama han consultado información política online comparado con el 57% de los seguidores online de Clinton y también han superado a los seguidores online de McCain (65% vs. 56%), habiendo éstos efectuado más contribuciones campaña, más firmas en peticiones online, más comentarios en blogs y más vídeos vistos que el resto del electorado (Smith y Rainie, 2008). Estas cifras pueden ser una clara consecuencia de que demócratas y en concreto los seguidores de Obama, mantienen una actitud “red” diferente a la que muestran los seguidores republicanos:

- mientras que los republicanos piensan que Internet está llena de “desinformación”, los demócratas se inclinan en mayor número a pensar que el uso de este medio les ayuda a sentirse más cercanos a su candidato,
- en general, los seguidores de Obama son más activos políticamente en la Red que los de Mc’Cain,
- la mayor inclinación al uso de Internet de los jóvenes votantes a favor de Obama y hacia los demócratas en general, entrega a éstos ventajas en el resultado de su campaña *online*.

Por otro lado, hemos visto como candidatos con mejores índices de popularidad en la Red no han visto reflejada tal ventaja en el resultado electoral

obtenido, como Antanas Mockus, candidato a la presidencia de Colombia por el Partido Verde era el preferido en redes sociales como Facebook¹ o Twitter² y, sin embargo, fue derrotado por su oponente Juan Manuel Santos, consiguiendo, no obstante, ser la segunda fuerza política del país con más de tres millones de votos recabados por su causa en las urnas y siendo éste su mejor resultado electoral hasta el momento. Y también podemos citar ejemplos en los que candidatos que no han desarrollado actividad web alguna, han resultado victoriosos en la contienda electoral, como es el caso del cercano Valcárcel en las pasadas Elecciones Autonómicas de la Región de Murcia, caso en el que se ha de tener en cuenta, entre otras variables, su posición de reelección, la cobertura recibida por los medios de comunicación tradicional, y las características ideológicas y demográficas del electorado.

En los últimos años, los medios y redes sociales se han convertido en un espacio del mayor interés para la colaboración social y compartir información entre sus usuarios. La ingente cantidad de contenido que circula en ellas ha atraído la atención de agencias de marketing que esperan conocer el comportamiento de los usuarios para ajustar sus campañas online y de investigadores en el campo de las redes sociales que buscan en sus análisis revelar los modelos de difusión entre grupos y nodos en dichas redes, e incluso demostrar como ese contenido social puede ser utilizado para predecir resultados en el mundo real (Asur y Huberman, 2010). Recientemente este tipo de análisis ha saltado al campo de la política, y especialmente, al análisis de campañas electorales en la red.

En el año 2010, un estudio del Gartner Group reveló las diez tecnologías estratégicas para el 2011. Entre ellas se citó el Análisis de Redes Sociales, que

¹ La página de Mockus en Facebook, <http://www.facebook.com/a.mockus>, cuenta con el doble de seguidores que la del vencedor Santos, <http://www.facebook.com/a.mockus#!/JMSantos.Presidente?v=wall&ref=ts>, (791,719 vs.369,128) (Consultado el 28/07/2010)

² Mockus cuenta con 65,855 seguidores. No se ha encontrado referencia sobre la cuenta en Twitter de Santos. (Consultado el 28/07/2010)

describió como el proceso de medir, analizar e interpretar los resultados de las interacciones y asociaciones entre las personas, temas e ideas. Incluye una serie de técnicas de análisis especializadas como el filtrado social, el análisis de redes sociales, el análisis de los sentimientos que circulan por ellas y el análisis de los medios sociales empleados. Las herramientas de análisis de redes sociales son útiles para examinar la estructura social y la interdependencia, así como las pautas de comportamiento de los individuos, grupos u organizaciones. El análisis de redes sociales incluye la recolección de datos de múltiples fuentes, identificando las relaciones, y evaluando el impacto, calidad o eficacia de una relación.

Teniendo en cuentas los principales medios sociales en Red en los que hoy en día, al menos en España, se articulan las Ciber campañas Electorales, podemos contemplar varios espacios para el análisis del impacto de una campaña en red en Facebook, Twitter, Google y la página Web de partido y/o de candidato.

En el campo de la ciencia política y concretamente en lo que se refiere a la capacidad de predecir un resultado electoral mediante el estudio de redes sociales, destaca el análisis de datos en Twitter tanto por ser una red extendida y reconocida como medio de expresión y difusión de opiniones, como por la facilidad que ofrece para capturar dichos datos gracias a su API de libre disposición y explotación que permite recolectar lo que se conoce como *public time-line*, esto es todos los tweets de sus usuarios salvo los de aquellos que mantienen un perfil privado que protege sus tweets del resto. Además de Twitter, también podemos encontrar estudios que relacionan el número de seguidores o fans de una página Facebook de un candidato con su resultado electoral, y otros que han utilizado el volumen de búsquedas realizadas en Google sobre una opción política determinada.

Debido a los positivos resultados obtenidos en estos primeros estudios se ha apelado a la idea de que lo que la gente está publicando en su blog, o buscando en Google o difundiendo en Facebook o Twitter, proporciona un reflejo del sentimiento colectivo, lo cual puede traducirse en tendencia hacia un punto determinado. En este

sentido, recientes estudios han anunciado que es posible predecir un resultado electoral mediante el análisis de este contenido social.

Concretamente, en “Predicting elections with Twitter: What 140 characters reveal about political sentiment”, de Tumasjan, Sprenger, Sandner y Welp (2010) se concluye que en las elecciones de 2009 en Alemania, Twitter puede ser utilizado como predictor del resultado electoral: “*The mere number of tweets mentioning a political party can be considered a plausible reflection of the vote share and its predictive power even comes close to traditional election polls.*” (Tumasjan, Sprenger y Sandner, 2010, p.183) Su teoría se basa en estudios anteriores y similares, como el de Véronis (2007) referido a las elecciones presidenciales de 2007 en Francia, en el que se expresa que el número de apariciones o de menciones de un candidato en la prensa puede ser un buen predictor del resultado de la contienda.

En “Form tweets to polls: linking test sentiment to public opinion time series”, O’Connor, Rammath, Bryen y Noah (2010) relacionaron la opinión pública en EEUU en el periodo 2008-2009 medida en las encuestas tradicionales con el sentimiento reflejado en los tweets (*sentiment analysis*). Sus resultados eran muy coincidentes en algunos casos (con una correlación del 80% en la mayoría) y menos en otros: para *presidential job approval* en 2009 los datos de Twitter replican los obtenidos en las encuestas y para *pesidential elections polls* in 2008 la correlación no fue tan significativa. Los autores sugieren que es necesario mejorar las herramientas de NLP (programación neurolingüística) para mejorar dichas estimaciones.

Tal y como hemos comentado antes, las características de privacidad de la red social Facebook dificultan la labor de análisis de interacción en la red. Si bien en el análisis de datos en Twitter hay que tener en cuenta o habría que tener en cuenta variables externas a la propia red, en Facebook, esta realidad es una obligación de entrada. Los análisis publicados hasta la fecha, siempre incorporan una serie bastante numerosa de variables a cada candidato, dado que en Facebook, salvo los perfiles de los propios candidatos y partidos (con datos como número de “like” o “supporter” o “seguidores”, comentarios y publicaciones) y su evolución en pre y campaña, es

prácticamente el único dato que puede obtenerse de manera automática por motivos derivados de las condiciones de privacidad de los usuarios de esta Red.

En varios trabajos de Willians y Gulati (2008, 2009), centrados en el análisis de Facebook en campaña electoral en EEUU, se concluye que el número de Fans constituye un indicador de la viabilidad del candidato y por tanto, del éxito electoral: *“Thus while Facebook supporter numbers would not be a useful predictor that foreshadows electoral victory or defeat, the most electable candidates do have more Facebook supporters.”* (Willians y Gulati, 2009:4) En su análisis incorporan una serie de variables para cada candidato, como su posición de peso en el partido, el hecho de que acudan a su reelección, que se encuentren en una campaña competitiva y las características de su electorado (alto nivel de educación, blancos, residentes en núcleos urbanos y jóvenes, son más propensos a participar en Facebook).

Google Trends es una herramienta de Google Labs que muestra los términos de búsqueda más populares del pasado reciente. Las gráficas de Google Trends representan con cuánta frecuencia se realiza una búsqueda particular en varias regiones del mundo y en varios idiomas. Por tanto, si buscamos en esta herramienta por una determinada opción política obtendremos la cantidad de veces que ha sido buscado en un periodo de tiempo concreto (campaña electoral), lo cual podría traducirse en que a mayor número de búsquedas de un candidato mayor posibilidades de éxito electoral que su oponente.

Lui, Metaxas y Mustafaraj (2011) en su análisis en Google Trends en las elecciones de 2008 y 2010 en EEUU, concluyen que Google Trends no es un buen predictor en ninguno de los dos procesos electorales analizados. Los autores argumentan que las causas de estas desviaciones pueden deberse por un lado al hecho de cuando un usuario busca por un candidato y no por su oponente, el resultado es que este último no alcanza los niveles mínimos de curiosidad y/visibilidad (permaneciendo en el vacío) lo que repercute más positivamente en el candidato buscado; por otro lado, que los electores pueden buscar por un candidato pero con el objetivo de encontrar información negativa lo que provoca ese margen de error con

los resultados electorales; y finalmente, que aquellos candidatos que ya ocupan un cargo político sobresaliente pueden ser buscados a causa de esta actividad y no tratando de encontrar información sobre su campaña.

¿Y el impacto de las páginas Web's de partidos o candidatos? En los estudios realizados sobre este tema, al igual que encurría con el análisis de Facebook, se incorporan una serie de variables relacionadas directamente con el “valor” o “posición” del candidato en la campaña.

La aparición de Howard Dean en 2003 ha sido señalada como *the coming of age* en el mundo de las campañas en Internet. Varios autores han estudiado el impacto de la página web de un candidato en su resultado electoral. D'Alessio³ en su análisis de las elecciones al Congreso de EEUU en 1996 ya concluía que contar con una cibercampaña proveía al candidato con 9,3 más votos de los previstos inicialmente, valores únicamente superados por la variable de la incumbencia, esto es, presentarse a la reelección, y por la posición más o menos relevante que ocupa dentro de su partido. Más recientemente, Gibson y McAllister (2005) estudiaron el efecto de las cibercampañas en el resultado electoral en las elecciones australianas de 2004, encontrando similares resultados: una cibercamapaña es un elemento primordial a la hora de asegurar la victoria electoral. El uso de una página Web proporciona un 4% adicional a la preferencia de voto inicial, lo cual supone más que todos los métodos tradicionales de campaña, excepto la incumbencia⁴.

Hemos visto hasta ahora como en el análisis de redes sociales como método de predicción electoral ha obtenido resultados dispares. Así, mientras que siguiendo los métodos tradicionales de encuestas, el margen de error no supera o debe superar

³ D'ALESSIO, David, W. (1997). “Use of the Web in the 1996 US Election”. Electoral Studies 16(4): 489-501. Mencionado por Gibson y McAllister (2005).

⁴ Del inglés “incumbency”. En política, candidato que se presenta a la reelección del cargo que ya ostenta (incumbent).

el 2 o 3%, en este tipo de análisis el margen de error puede llegar a alcanzar, e incluso superar, el 10%.

Daniel Gayo-Avello (2011), desde la Universidad de Oviedo, advierte sobre esta facilidad de convertir los medios sociales en el futuro *Literary Digest*⁵ en su paper “A warning against converting Social Media into the next *Literary Digest*” en el que su análisis electoral en Twitter, llevado a cabo replicando métodos que en otros investigadores habían obtenido resultados favorables⁶, no obtuvo coincidencias suficientes para afirmar que Twitter puede ser empleado como herramienta de predicción electoral.

La predicción de mercados basada en las apuestas, han resultado ser una nueva fuente de pronóstico electoral. Varios estudios académicos han confirmado el valor de los mercados de predicción en el pronóstico de próximos eventos, para los resultados electorales, demostrando su eficacia desde las elecciones de finales del XIX y las primeras elecciones del siglo XX. Examinando 74 batallas electorales en los últimos 130 días de la campaña 2008 en EEUU y recolectando datos de sitios Web de como PollingReport.com, Pollster.com, RealClearPolitics.com e Intrade.com, David Rothschild exploró la precisión de las previsiones derivadas de dos tipos diferentes de datos: las encuestas y los mercados de predicción. Las conclusiones de su estudio confirman los resultados logrados por estudios similares, como el de Rhode y Strumpf en 2004: *In 2008, FiveThirtyEight, a debiased poll-based forecast, offered to the general public a more accurate forecast than raw poll*

⁵ La publicación The Literary Digest llevó a cabo un "sondeo" con el ánimo de avanzar el resultado de las elecciones presidenciales de 1936. El sondeo mostró que el gobernador republicano de Kansas, Alf Landon, probablemente sería el ganador abrumadoramente. En noviembre, Landon lleva sólo Vermont y Maine, y el presidente Franklin Delano Roosevelt ganó en los estados 46 estados restantes de entonces. La revista fue desacreditada por completo a causa de la encuesta y desapareció pronto. En retrospectiva, las técnicas de votación empleadas por la revista fueron las culpables: los encuestados eran en su mayoría lectores de su revista, con un gran nivel adquisitivo y los datos de los encuestados se extrajeron de bases de datos de poseedores de automóviles y de teléfonos en sus hogares. http://en.wikipedia.org/wiki/The_Literary_Digest. (Consultado el 11/02/21)

⁶ Si bien incorporó algunas variaciones, como la elección de la herramienta para investigar el *Sentiment Analysis*.

numbers or raw prediction market. But, the analysis here shows that were Intrade's prices debiased, they would have provided a more accurate forecast and more valuable information than the best poll-based forecasts currently available, especially early in the cycle and in uncertain races." (Rothschild, 2009:913)

Con similar metodología, Nate Silver un estadístico del béisbol que se inició el análisis de las encuestas políticas, presentó en marzo de 2008 su sitio *FiveThirtyEight.com*, en el que utiliza su propia fórmula para predecir los resultados federales y estatales de las elecciones presidenciales de 2008 en EEUU, prediciendo la victoria de Barack Obama.

A tenor de las conclusiones vertidas por los estudios mencionados, podría adelantarse que no siempre es posible predecir elecciones a partir de los datos extraídos de la participación de los individuos en las cibercampañas a través del uso de las redes sociales. Por una parte, los resultados obtenidos en un proceso, no son directamente extrapolables a otro, tanto porque los sistemas electorales son diferentes como porque los espacios en los que se producen presentan características demográficas bien diferentes. Por otra, los métodos de correlación que se emplean son también diferentes, siendo de unos casos los datos comparados con resultados electorales y en otros con encuestas preelectorales. Y por último, las herramientas y métodos de análisis, novedosos hasta el momento, deben ser testados y continuamente mejorados para asegurar el veraz resultado de los datos obtenidos. Además, hemos de tener en cuenta que no todas las franjas de edad son usuarios activos en estas redes sociales, siendo los mayores de 50 años los menos propensos a participar y compartir así sus opiniones y tendencias de voto.

No obstante, se abre un campo de especial interés y relevancia de cara al futuro, en el que tal vez llegue a ser posible sustituir a las encuestas, que hasta hoy han sido una fuente fiable de la tendencia electoral. Así, el objetivo de esta Tesis Doctoral y los resultados obtenidos en ella aportarán nuevos estudios de caso en esta reciente disciplina de la predicción electoral mediante análisis de Redes Sociales,

monitorizando la actividad en Twitter de cuatro procesos electorales sucedidos en España, y ayudarán a refutar otras hipótesis de partida.

A priori, la hipótesis de partida se refiere a la presunción de que aquellos partidos y/o candidatos que muestren mejores valores en el seguimiento y participación de su campaña online, obtendrán mejores resultados electorales. A este tenor y como consecuencia, la misma hipótesis resultaría en la afirmación de que el análisis de Redes Sociales en Internet serviría como herramienta fiable para predecir resultados electorales.

El presente trabajo de investigación está organizado en cuatro capítulos, cuyos contenidos se refieren a continuación.

En el Capítulo 1, El comportamiento de las Redes Sociales, dado el objetivo global de este trabajo de investigación, se trata de conocer las características y reglas de una red social así como descubrir los comportamientos que los individuos desarrollan en las Redes Sociales en Internet, son aspectos fundamentales a tratar e indispensables tener en cuenta y conocer para entender tanto el marco teórico en el que la investigación se desenvuelve como los motivos que la han llevado a elegir el tema objetivo como el análisis de los datos capturados.

En este primer capítulo intentaremos dilucidar si aspectos como la difusión boca a boca, la capacidad de influencia, los vínculos débiles, la homofilia y, en suma, el comportamiento de los individuos y de sus ideas en las redes sociales en Internet responden a los modelos tradicionales que se dan en las redes sociales “físicas”, y en caso no coincidentes, desvelar los rasgos o detalles que marcan las desviaciones encontradas.

La primera parte del mismo está dedicada al conocimiento básico de los mecanismos de difusión y contagio que se dan en las redes sociales. A continuación, se centra en las particularidades de Internet como red, y concretamente en las propiedades de las Redes Sociales Facebook y Twitter. Se eligen éstas y no otras

para su estudio, dada su gran aceptación entre la población española, entorno en el que se despliega la presente investigación.

Termina con un epígrafe centrado en como ocurre el contagio y concretamente el contagio político, dentro de una Red Social, física y en Internet.

En el Capítulo 2, Marco Teórico, se han incluido y revisado los temas que se han considerado relacionados con distintos aspectos que atañen al objeto de esta investigación. En su conjunto, se trata de una serie de datos estadísticos y estudios de investigación con objeto similar al que nos hemos propuesto. Alguno de estos estudios han sido el origen y la inspiración de esta Tesis Doctoral, y otros han servido para extraer piezas claves de conocimiento probado que se incorporarán a la metodología empleada.

Los temas en los que nos hemos fijado son, por orden de aparición en el capítulo, los siguientes:

- “El auge de la política en Internet”, epígrafe en el que se pone de manifiesto tanto el incremento del uso de Internet por parte de los usuarios para informarse sobre diversas opciones políticas como la intensificación de la presencia de candidatos y partidos en la Red, sobre todo para divulgación y promoción de sus campañas electorales.
- “Antecedentes del uso del análisis de Redes Sociales como herramienta de predicción”, apartado dedicado a la exposición de estudios realizados mediante el análisis de la actividad de los usuarios en Internet con el objeto de predecir las tendencias de opinión pública.
- Tres epígrafes - referidos a “Facebook, Google y Twitter como herramientas de predicción electoral”-, en los que recorren los estudios publicados por diferentes autores que han vigilado distintas contiendas

electorales a través de la monitorización de los datos que los usuarios de estas redes vertían en cada una de ellas⁷.

- “El peso de las Ciber campañas en el éxito electoral”, en el que evidencia el mayor o menor peso que la dotación de una Ciber campaña puede aportar de cara a la captación de votos y por tanto, a la mejora del resultado electoral.
- “Encuestas preelectorales. Eficacia y desviación”, dedicado a constatar la realidad de las encuestas preelectorales como métodos fiables de predicción de los resultados electorales.
- “Grado de penetración de las Redes Sociales en la población española”, destinado a la obtención de datos característicos de los usuarios de las Redes Sociales en España

A priori, y continuando con la hipótesis de partida, afirmaríamos que si las citadas Redes Sociales han obtenido resultados positivos en los estudios de tendencias de opinión pública, demostrando que es posible su análisis para una toma instantánea de la realidad, y dado que está comprobado que disponer de una campaña online mejora los resultados electorales de aquellos que hacen uso de ella, continuamos remarcando que de igual forma, dichos análisis pudiera servir para predecir un posible resultado electoral. Además, sería posible establecer una comparación su eficacia con la demostrada por las encuestas de opinión tradicionales, que si bien vienen demostrando su profesionalidad y valor predictivo, también han dado muestras de que sus métodos no son infalibles. Valga citar como ejemplo fallido, cercano temporal y espacialmente, las encuestas referidas a las recientes Elecciones al Parlamento Andaluz 2012, en el que apenas se acercaron a la predicción de la futura composición del mismo.

⁷ Si bien Google no es en sí misma una Red Social como Facebook, Tuenti, Badoo, etc., es un potente buscador Web, utilizado masivamente en todo el mundo, que basa el acierto de los resultados de las búsquedas en un algoritmo propio que vigila y analiza los gustos y predilecciones de sus millones de usuarios, y cuyas herramientas de interpretación de estas tendencias sociales pueden ser utilizadas a título particular, para extraer información valiosa de marcado carácter social.

En el Capítulo 3, Metodología, se expone el método empleado para el análisis realizado sobre cuatro campañas electorales en España, acontecidas en los años 2010 y 2011, y que concretamente se refiere al análisis de los datos capturados en la red Twitter durante el periodo de campaña electoral (15 jornadas previas a la jornada electoral):

- Elecciones al Parlamento de Cataluña 2010
- Elecciones Autonómicas de la Comunidad de Madrid 2011
- Elecciones Autonómicas de la Región de Murcia 2011
- Elecciones Generales 2011

La selección de estos cuatro procesos electorales viene motivada por las siguientes razones de peso:

- La novedad de la propia Red Social (Twitter), ya que antes del año 2010 su uso en España era bastante reducido en lo que se refiere a número de participantes en la misma, lo que supondría que la sociedad española apenas se encontraría representada.
- Las primeras investigaciones basadas en el análisis de datos en Twitter datan del año 2010, por lo que es a partir del conocimiento de su posibilidad cuando iniciamos la captura de datos del primer proceso electoral referenciado.
- Se monitorizan las elecciones que tuvieron lugar en el transcurso de un año. Ampliar dicho periodo dificultaría el establecimiento de elementos comunes de comparación entre los objetos observados, dado que entendemos que la población de estudio soportaría cambios estructurales relevantes como un incremento significativo del número de usuarios Twitter.
- Al tratar de procesos electorales que suceden en distintos ámbitos geográficos, siendo uno de ellos de carácter nacional y el resto de ámbito autonómico, los resultados de sus análisis podrán arrojar conclusiones que indiquen las posibles diferencias entre el comportamiento Red en una

elección atendiendo a las particularidades del entorno en que ésta se produce.

Fijándonos en métodos utilizados en varios de los trabajos mencionados a lo largo de esta obra, se ha decidido seguir como base inicial la técnica empleada en el publicado por Tumasjan, Sprenger, Sandner y Welp (2010) en el que sus autores concluyen que el número de *tweets* en los que se menciona (actividad Twitter conocida como “menciones”) a una u otra opción política puede considerarse como un reflejo válido y previsible de la intención de voto, y por ende, del resultado electoral, acercándose a los métodos predictivos obtenidos por las encuestas tradicionales.

Se ha desestimado el “análisis sentimental” o *sentiment analysis* por varias razones. La primera y no desdeñable, es la complejidad que conlleva su estudio puesto que las herramientas de análisis disponibles tratan de extraer un sentido comunicativo exacto partiendo de la captura aislada de términos o pares de términos fuera de su contexto, y la segunda, y no menos importante, es la constatación por Asur y Huberman (2010) de que este tipo de análisis no incorpora mejoras en comparación con el análisis único de los tweets: “*We find that the sentiments do provide improvements, although they are not as important as the rate of tweets themselves.*” (Asur, y Huberman, 2010:7). Por otro lado y contrariamente a lo que sucede en otros ámbitos, mientras que se venía considerando que los adjetivos afectivos cumplían un papel definitivo en la clasificación de los sentimientos, en el dominio de la política, estas palabras no son los mayores indicadores del tipo de la información que se vierte, siendo las denominaciones (de partidos y candidatos) modos de expresión política válidos en sí mismos:

“Furthermore, affective adjectives are not the most informative political opinion indicators. Instead, the choice of topics, as reflected in neutral nouns, is an important mode of political opinion expression by itself. Our manual annotation results

demonstrate that a significant number of political opinions are expressed in neutral tones.” (Yu, Kaufman y Diermeier, 2008)

Se ha dividido en dos apartados principales, que explican el método empleado para la fijación de la metodología y el procedimiento seguido en la captura de datos en esta red.

En el Capítulo 4, Resultados de la Investigación, se muestran los resultados finales de la investigación obtenidos para cada una de las elecciones monitorizadas. En todos los casos, se han sumado de forma independiente todas y cada una de las actividades twitter, tomadas como variables, que una determinada opción política ha recibido en esta Red durante la campaña electoral, señalando las fechas de captura en cada caso. Se indica, por tanto, para cada una de las variables establecidas los *strings* o cadenas de texto que se han tomado como referencia para realizar el cómputo global, esto es, se han sumado aquellos *tweets* en los que dichas cadenas de texto aparecían.

A priori, la hipótesis de partida puede resultar simple: cuántas más veces es mencionada una determinada opción política por los usuarios de Twitter, más votos obtendría en la jornada electoral. Además, en un cálculo porcentual de las menciones recibidas por todas y cada una de ellas, el porcentaje que obtenga cada una de ellas se correspondería con el porcentaje de votos conseguidos en la elección.

Para la primera y segunda línea de investigación, los análisis que se han llevado a cabo en cada una de las elecciones y cuyos resultados finales se incorporan en sendas tablas, son los siguientes:

- Resultado Twitter
- Resultado Twitter Sólo Menciones
- Resultado Twitter Candidato
- Resultado Twitter Candidato Sólo Menciones
- Resultado Twitter Partido

- Resultado Twitter Partido Sólo Menciones

La introducción de una diferenciación según se tengan en cuenta o “sólo menciones” o no (se suman a las menciones todas las actividades que procura Twitter, como número de seguidores, *hashtag*, *RT* y *reply*), viene impuesta por la decisión de continuar el método empleado por los citados autores Tumasjan, Sprenger, Sandner y Welp (2010), y a su vez, ampliarlo con métodos similares pero no idénticos.

Dado que en Twitter es posible referenciar tanto a candidatos como a partidos, se han efectuado diferentes cálculos, bien sea teniendo en cuenta el total de la actividad Twitter, o la hallada para candidatos y partidos separadamente.

Además, con el objeto de seguir la metodología empleada por Tumasjan et al., (2010), en cada uno de los casos, se ha calculado el resultado teniendo en cuenta las menciones de forma única y evitando el resto de actividad Twitter, como *hashtag*, *RT*, *reply* o *tweets* publicados por sus protagonistas.

Para completar el alcance y conclusiones de esta investigación se ha establecido una comparación entre los resultados que la vigilancia de esta Red y los que arrojan las encuestas de opinión tradicionales en los cuatro escenarios electorales analizados. De este modo, se presentan en sendas tablas, las diferencias que ambos métodos han obtenido en relación con los resultados electorales definitivos.

Para completar el análisis, se ha llevado a cabo una aproximación a una tercera línea de investigación, referida ésta a un cambio en el periodo de datos capturados, reduciéndose de quince a siete los días de recogida de datos, siendo estas jornadas las siete justamente anteriores a la celebración electoral. Se hace patente que en esta tercera línea de investigación intentamos una aproximación a diversos estudios anteriores y de diversas disciplinas -Desart y Holbrook (2003) y Jungherr, J’Urgens y Schoen (2011), en los que se asegura que las fechas en las que los datos son recolectados, influyen positiva o negativamente en los resultados, adquiriendo,

según los primeros, mayor certeza cuanto más próximos se encuentren de la jornada electoral, e incrementa la desviación según sugieren los segundos⁸.

Finalmente, volvemos a retomar el peso que en este tipo de investigaciones puede llegar a alcanzar el tamaño de la muestra. Debe entenderse por muestra, no sólo el porcentaje de población con participación en Twitter sino, y de forma esencial, el volumen de la actividad de ésta. Es constatable que las diferencias en cantidad entre una elección general y una elección regional, son considerables, lo cual nos anima a considerar que pueda ser necesaria una cantidad de datos o *tweets* proporcional al universo poblacional de dicha elección, tal y como se concluyó en el estudio de Teewtminster (2010) en el que se afirmaba que la correlación es mayor cuando el volumen de *tweets* de una elección es considerable, y a más, las características demográficas no son determinantes en estos casos.

Continuando con el ánimo de encontrar una relación de tipo causal entre el MAE⁹ que arroja los resultados de Twitter y los datos referidos a participación y población, también se contabilizan las relaciones que se establecen entre “usuarios únicos¹⁰” y “usuarios twitter”, y entre aquellos y el “volumen o número de tweets”, para comprobar si mantienen una relación directa con la fiabilidad del carácter predictivo obtenido en Twitter.

A priori, se parte de la hipótesis de que el número de “usuarios únicos” es determinante para el valor de los datos capturados y el consiguiente buen resultado de la investigación: cada uno de dichos “usuarios únicos” vendría a corresponderse con un miembro del electorado y posible votante que expresa su tendencia o preferencia de voto en Twitter, de forma que a mayor número de éstos mayor tamaño de la muestra encuestada y, por tanto, mayor fiabilidad de la misma.

⁸ Podrá apreciarse que sólo se incluyen los resultados de dos de las cuatro elecciones monitorizadas. Los procesos electorales elegidos han sido las Elecciones Generales 2011 y las Elecciones Autonómicas a la Comunidad de Madrid, por motivos justificados que se enuncian en dicho apartado.

⁹ *Mean Average Percentage Error*.

¹⁰ Cifra de individuos diferentes que participaron en una determinada campaña electoral en Twitter.

Por último, en las Conclusiones, se refutan las hipótesis planteadas en la investigación intentando obtener en base a los resultados obtenidos y en relación a las afirmaciones del resto de autores consultados, nuevas líneas de aproximación metodológica que adviertan de las debilidades del entorno y de las herramientas del análisis, y propongan situaciones deseadas para optimizar la rentabilidad y eficacia final de futuros estudios electorales en lo que análisis de Redes Sociales se refiere.

Capítulo 1

El comportamiento de las Redes Sociales

2. Capítulo I: El comportamiento de las Redes Sociales

2.1. Introducción

El creciente interés por el efecto de Internet y de sus Redes Sociales en el campo de la comunicación política es un hecho patente. Políticos y asesores de comunicación política se ven prácticamente obligados a desarrollar campañas en un medio un tanto desconocido por su novedad y en el que podría ser factible que las reglas aplicadas a medios y entornos tradicionales, no funcionen como esperaban.

Las redes sociales “físicas” se vienen estudiando desde hace décadas. De su análisis se desprende cómo actuar en ellas para conseguir un objetivo, que en el caso que nos ocupa no es otro que difundir un mensaje y contagiar a otros para que voten por el candidato de nuestra elección. Pero existen indicios, o al menos, cierta incertidumbre provocada por el desconocimiento del nuevo canal, que señalan que en Internet las reglas de transmisión, contagio e influencia no replican las ya conocidas.

Es por ello y por el objetivo global de este trabajo de investigación, que conocer las características y reglas de una red social así como descubrir los comportamientos que los individuos desarrollan en las Redes Sociales en Internet, son aspectos fundamentales a tratar e indispensables tener en cuenta y conocer para entender, tanto el marco teórico en el que la investigación se desenvuelve, como los motivos que la han llevado a elegir el tema objeto como el análisis de los datos capturados.

En este primer capítulo intentaremos dilucidar si aspectos como la difusión boca a boca, la capacidad de influencia, los vínculos débiles, la homofilia y, en suma, el comportamiento de los individuos y de sus ideas en las redes sociales en Internet responden a los modelos tradicionales que se dan en las redes sociales “físicas”, y en

caso no coincidentes, desvelar los rasgos o detalles que marcan las desviaciones encontradas.

La primera parte del mismo está dedicada al conocimiento básico de los mecanismos de difusión y contagio que se dan en las redes sociales. A continuación, se centra en las particularidades de Internet como red, y concretamente en las propiedades de las Redes Sociales Facebook y Twitter, elegidas por su gran aceptación entre la población de España –entorno en el que se despliega la investigación. Nos referimos a Facebook y Twitter de forma especial, además de por caracterizarse por ser dos de las Redes Sociales más empleadas por la sociedad de nuestro entorno, porque son espacios Web destacados en los que nuestros políticos despliegan sus cibercampañas. Termina con un epígrafe centrado en cómo ocurre el contagio político dentro de una Red Social, física y en Internet.

2.2. Conociendo las Redes Sociales

2.2.1. El concepto de Red Social

Una de las formas más sencillas de definir una red social la encontramos en las palabras de Nicholas A. Christakis y James H. Fowler cuando la identifican como *“un conjunto organizado de personas formado por dos tipos de elementos: seres humanos y conexiones entre ellos”*. (Christakis y Fowler, 2010, p.27)

El antropólogo J. A. Barnes, en el resultado de su estudio sobre los habitantes de las islas noruegas, la definía de la siguiente manera: *“Cada persona está, por así decirlo, en contacto con cierto número de otras personas, algunas de las cuales están en contacto entre sí y otras no. Creo conveniente denominar red a un campo social de este tipo. La imagen que tengo es la de una red de puntos los cuales algunos están unidos por líneas. Los puntos de esta imagen unas veces serán personas y otras grupos, y las líneas indicarían quiénes interactúan entre sí.”* (Barnes, 1954, p.43)

Barnes, fue el primero en introducir el concepto de red desde un punto de vista analítico, otorgando importancia en estas estructuras, no sólo a las propiedades o características de los componentes del grupo - *puntos* -, sino también a las relaciones - *líneas* - que se establecen entre ellos.

En la teoría matemática de grafos¹¹, utilizada en el análisis de redes sociales, “*se llama red a una serie de puntos vinculados por una serie de relaciones que cumplen determinadas propiedades. (...) Es un conjunto de relaciones en el cual las líneas que conectan los diferentes puntos tienen una valor concreto, sea éste numérico o no.*” (Santos, 1989, p.139)

En su libro sobre la sociedad red, Manuel Castells coincide con los postulados anteriores, asignando valor a los nodos - puntos - y las conexiones que se establecen entre ellos, afirmando que “*Una red es un conjunto de nodos interconectados. Una red no posee ningún centro, sólo nodos. Los nodos pueden tener mayor o menor relevancia para el conjunto de la red (...) La red es la unidad, no el nodo.*” (Castells, 2006, p.27)

El concepto de red social está relacionado en muchos aspectos con el de sociograma, representación gráfica de una estructura grupal. Su técnica, la sociometría, fue desarrollada en los años 30 por Jacob Levy Moreno¹² y definida en su libro *Fundamentos de la sociometría* “como el estudio matemático de las propiedades psicológicas de las poblaciones, la técnica experimental de aplicación de métodos cuantitativos a las relaciones humanas, la exploración de la evolución y organización de grupos y la posición de los individuos dentro de ellos, así como la ciencia de la organización de los grupos.” (Zapata, 2006:171). La radiografía grupal busca obtener de manera gráfica, mediante la observación y contextualización, las

¹¹ El trabajo de Leonhard Euler, en 1736, sobre el problema de los puentes de Königsberg es considerado el primer resultado de la teoría de grafos. http://es.wikipedia.org/wiki/Problema_de_los_puentes_de_K%C3%B6nigsberg. (Consultado el 23/08/2011)

¹² Dr. Jacob (Jakob) Levy Moreno (1889-1974) fue un reconocido psiquiatra, teórico y educador. Es el fundador del Psicodrama, la Sociometría y uno de los pioneros de la Psicoterapia Grupal.

distintas relaciones entre sujetos que conforman un grupo, poniendo así de manifiesto los lazos de influencia y de preferencia que existen en el mismo. Este tipo de relaciones no son necesariamente formales, en la mayoría de los casos son informales.

Podemos concluir y proponer como definición, que una red social está formada por las personas (nodos o vértices) y las relaciones que se establecen entre ellas (aristas o enlaces), siendo estas últimas las que toman fuerza en el análisis de redes sociales, dado que, como hemos apuntado, no se concede sólo importancia a los atributos particulares de los actores -clase social, profesión o grupo étnico-, sino que se añaden como atributos los vínculos que relacionan a unos con otros, y se toma en cuenta cómo lo que discurre por la red afecta a los implicados en la misma -contagio-.

2.2.2. Propiedades de las redes sociales

La principal característica que se desprende al estudiar la estructura interna de las redes sociales (por ejemplo, de un grupo de estudiantes de una Universidad), es la posición que ocupamos en ellas. Podemos situarnos en el centro o en la periferia de nuestra red. Nuestra posición repercute en la capacidad que tenemos de influir en el resto de actores de la red a la vez que se relaciona con los conceptos de autonomía e independencia: a posiciones más centrales, mayor capacidad de influencia y grado de autonomía para actuar sin contar o necesitar al resto. Dicho de otra manera, el poder es consecuencia de la posición en la red: *“las posiciones de los actores en una red social determinan la “estructura de oportunidad” de un actor respecto a la facilidad de acceder a los recursos de otros actores en la red.”* (Santos, 1989, p.141)

La centralidad es un atributo estructural de los nodos de una red que se relaciona con una mayor posición central (y menos periférica) que adquieren los

nodos en ella¹³. Fue introducida por el psicólogo social Alexander Bavelas en 1948. En la serie de experimentos empíricos que realizó a finales de los cuarenta en el *Group Networks Laboratory* del M.I.T se concluyó que “(...) *la centralidad estaba relacionada con la eficiencia del grupo a la hora de resolver problemas, con la percepción del liderazgo y con la satisfacción personal de los miembros del mismo.*” (Freeman, 1978, p.1)

Los valores que se emplean para medirla son: (Quiroga, Martí, Jariego y Molina, 2005)

- Grado nodal o *degree* (puede referenciarse como “conectividad). El grado nodal o rango es el número de lazos directos de un actor (o nodo), es decir con cuántos otros nodos se encuentra directamente conectado. Puede interpretarse cómo la oportunidad de influir o ser influido por otros nodos.
- Grado de Intermediación o *betweenness* (carga). El grado de intermediación indica la frecuencia con que aparece un nodo en el tramo más corto que conecta a otros dos. Es decir, muestra cuando una persona es intermediaria entre otras dos personas del mismo grupo que no se conocen entre sí (lo que podríamos denominar “persona puente”). La frecuencia con la que un nodo aparece en el camino más corto que conecta otros dos nodos. Indica el control del flujo por la red.
- Grado de Cercanía o *closeness* (accesibilidad). El grado de cercanía indica la cercanía de un nodo respecto del resto de la red. Representa la capacidad que tiene un nodo de alcanzar a los demás. Puede interpretarse como el tiempo que toma lo que fluye por la red en llegar de un nodo a otro.

¹³ Dado su peso en el análisis de las redes sociales, dedicaremos un epígrafe al concepto de centralidad.

Cuando nuestros amigos y nuestra familia se conectan mejor aumenta nuestro nivel de conexión con el conjunto de la red social. Nosotros decimos que esto nos “centra” porque contar con amigos bien conectados nos aleja literalmente de los márgenes hacia el centro de la red social.” (Christakis y Fowler, 2010, p.27)

Desde el punto de vista morfológico, la forma de una red es una de sus propiedades básicas e inmutables. Con independencia del método que se emplea para su visualización, el entramado de conexiones que contiene y la determina continúa siendo el mismo; la localización de un nodo en la red no cambia.

Localización o anclaje, junto con accesibilidad, densidad y rango, constituyen los elementos morfológicos que J. Clyde Mitchell distingue en las redes sociales: (Santos, 1989, pp.141-143)

- Anclaje. El punto de origen o actor inicial desde el que se procede a la ramificación de la red. Una red tiene que estar anclada en un punto de referencia
- Accesibilidad. La fuerza con la que un actor está influenciado por otros. La accesibilidad en la red puede medirse con la proporción de actores que pueden contactar con un actor y el número de vínculos necesarios para contactar con un actor concreto.
- Densidad. Depende del número de vínculos que exista dentro de la red. Así, una red en la que todos sus nodos están conectados tiene una densidad máxima. Por lo general, en una misma red, pueden observarse zonas de gran densidad y zonas de baja densidad.
- Rango. Podría definirse como el número de personas que están vinculadas directamente – sin intermediarios- con un individuo.

Además de contribuir al estudio morfológico de una red social, Mitchell¹⁴ definió las cualidades que determinan el tipo de relación que se establece dentro de una red concreta: (Santos, 1989, pp.143-145)

- Contenido. Referido al contenido de un vínculo en la red, lo que fluye por ella. Las diferencias de contenido se aprecian en redes superpuestas (por ejemplo, vínculos de parentesco y vínculos ocupacionales)
- Direccionalidad. Existen relaciones recíprocas (familiares) y otras de carácter más unidireccional (jefe-empleado). Según la dirección de la interacción existirá o no influencia de un actor sobre otro.
- Duración. Es el periodo de vida de una red social, durante el que puede expandirse o contraerse
- Intensidad. Puede definirse como el grado de implicación de sus actores, la mayor o menor incidencia que sobre el comportamiento de un actor tienen los otros.
- Frecuencia. Necesidad de una repetición del contacto entre los actores vinculados. Su importancia en el análisis de redes es marginal.

Una última característica de las relaciones en las redes sociales es su transitividad. La transitividad viene dada por el hecho de que todos los actores de una red se conecten o relacionen entre sí. Si A conoce a B y B conoce a C y C conoce a A, nos encontramos con una relación transitiva. Esta propiedad depende de la fuerza de los vínculos y se encuentra estrechamente relacionado con la teoría de “La fuerza de los vínculos débiles” expuesta por Mark S. Granovetter en 1973¹⁵. Christakis y Fowler (2010) coinciden con él al expresar que “*Las relaciones de alta transitividad suelen estar profundamente incrustadas dentro de un solo grupo, las de baja transitividad suelen afectar a personas de varios grupos distintos que no se conocen*

¹⁴ MITCHELL, J. Clyde, (1969) "The concept and use of social networks", en el libro editado por él mismo, *Social Networks in urban situations*, Manchester, Manchester University Press, pp. 1-50. Recogido en Santos (1989)

¹⁵ Dado su peso en el análisis de las redes sociales, dedicaremos un epígrafe al concepto de centralidad

entre sí, con lo cual, es más probable que actúen como puente entre los distintos grupos.” (Christakis, 2010, p.32)

2.2.3. Las reglas de la red

El valor de una red quizás resida en el hecho probado de que los vínculos pueden hacer que la suma del todo sea mayor que la suma de sus partes. Las conexiones que se establecen y el contagio de lo que fluye por la red son los dos aspectos fundamentales que lo hacen posible.

Para comprender por qué existen y cómo funcionan es necesario comprender las normas que rigen sus conexiones y la forma en que se produce el contagio. N. Christakis propone la siguiente lista de reglas en la red: (Christakis y Fowler, 2010, pp.30-39)

- Somos nosotros quienes damos forma a nuestra red.
 - Tendencia consciente o inconsciente a la homofilia, esto es, a asociarnos con personas que se parecen a nosotros
 - Elección de la estructura de la red. Decidimos a cuántas personas estamos conectados, modificamos la forma en que nuestra familia y amigos están conectados, y ejercemos cierto control sobre el lugar que ocupamos en ellas, central o periférico.
- Nuestra red nos da forma a nosotros. Nuestra localización, centralizada o periférica, en la red nos influye.
- Nuestros amigos nos influyen. Lo que fluye por las conexiones nos afecta. Cada uno de los vínculos propicia oportunidades de influir y de recibir influencias.

- Los amigos de nuestros amigos también nos influyen. Difusión hiperdiádica o la tendencia de los efectos de pasar de persona en persona más allá de los vínculos sociales directos de un individuo.¹⁶
- La red tiene vida propia. Propiedades emergentes de carácter grupal, atributos del todo que surgen de la interacción y conexión de las partes.¹⁷

2.3. Teorías de redes sociales

2.3.1. La importancia de la centralidad

Muchos son los estudios académicos que desde hace décadas han demostrado el papel relevante que juega la centralidad en una red. Desde los trabajos de B. S. Cohn y M. Marriot que en 1958 aplicaron el concepto de centralidad para explicar la integración política en un contexto tan grande y heterogéneo como la India, F. R. Pitts que en 1965 reconstruyó la red fluvial de la Rusia medieval para intentar comprender el posterior desarrollo de Moscú como epicentro de una geografía tan extensa, hasta Czepiel que en 1974 aplicó el concepto de centralidad a la difusión de las innovaciones en la industria del acero, concluyendo que la centralidad era un factor determinante en la adopción temprana de nuevos comportamientos –nuevos procesos de fundición- *“Czepiel (1974) applied the centrality measures in marketing in a study of the diffusion of a major technological innovation in industrial settings. He found that centrality was associated with early adoption behavior, although the relationship was not significant.”* (Vilpponen, Winter y Sundqvist, 2006, p.67)

H. J. Leavitt promulgó en 1951 que ocupar una posición central en una red supone que ese nodo no depende de otros como intermediarios o transmisores y que,

¹⁶ Véase como referencia el experimento del psicólogo Stanley Milgram en una acera de New York en 1968. Los peatones imitan la acción previamente programada de un grupo de personas que miran hacia una ventana. En Milgram, S., Bickman, L. y Berkowitz, L. (1969) “Note on the Drawing Paper of Crowds of Different Size”, *Journal of Personality and Social Psychology*, pp. 79-82.

¹⁷ Es el caso de los modelos matemáticos de las bandadas de pájaros y bancos de peces que se mueven al unísono. No hay control central del movimiento del grupo. El grupo manifiesta un tipo de inteligencia colectiva.

además, los conceptos de centralidad e independencia - la independencia de un punto está determinada por su cercanía a todos los demás puntos del grafo- eran intercambiables. (Granovetter, 1973, p.18)

Posteriormente otros autores han avalado estas hipótesis llegando a constatar que la cercanía es determinante para la centralidad: *“En un estudio anterior, Bavelas (1948) había sugerido otra lógica alternativa para construir una noción de la centralidad basada en la cercanía. Apuntaba la idea de que un mensaje originado en el punto más central de una red se expandiría a través de toda ella en un tiempo mínimo. Beauchamp (1965) llevó esta idea más lejos al hablar del uso de la cercanía para diseñar organizaciones de una “óptima... eficiencia” en sus comunicaciones. Hakimi (1965) y Sabidussi (1966) generalizaron esta idea al definir el punto más central de una red como aquel para el que el tiempo o el coste de la comunicación con los demás es mínimo.”* (Granovetter, 1973, p.18)

Las teorías sobre el poder de la centralidad en la estructura de una red social son de gran importancia ya que pueden determinar la capacidad que un actor central adquiere para contagiar lo que transmite, y sustentar la idea de que aquellos nodos bien posicionados –centrales- demuestran cierta supremacía sobre las ideas o comunicación del resto y que, por tanto, alcanzar los objetivos que se planteen – propios o de terceros- les resulta mucho más fácil que al resto.

Una prueba de campo concreta que consolida la teoría del peso de la centralidad en una red social determinada la encontramos en el funcionamiento de la aprobación de leyes en el Congreso de Estados Unidos. Aunque en el capítulo “El político mejor conectado”, Christakis y Fowler, se refieren a este efecto como “la fortaleza de las relaciones entre representantes políticos”, podemos considerar que a tenor de los supuestos expuestos, las conclusiones de su trabajo de investigación sobre los presupuestos de ley que pasan por el Congreso corroboran la relevancia de la centralidad: *“En otras palabras, si la ley es propuesta por una persona situada en el centro de la red, será aprobada, pero si la misma ley la propone alguien de la periferia, no será aprobada.”* (Christakis y Fowler, 2010, p.212) También

descubrieron que las personas que trabajan juntas tienden a estar más cercanas, incluso perteneciendo a partidos políticos diferentes y que, a la inversa, se identificaron enemistades personales entre representantes de ideologías similares.

La centralidad también influye en la percepción de liderazgo, la rapidez y eficiencia de una red en resolver sus problemas y la satisfacción de sus miembros, como demostró Leavitt en 1951. (Granovetter, 1973, p.23). Es decir que por sus efectos en el control, la independencia y la actividad, la centralidad afecta a los procesos grupales.

2.3.2. El mundo pequeño de Milgram (Seis grados de separación)

El psicólogo estadounidense Stanley Milgram realizó tres experimentos (1967, 1969 y 1970) que demostraron la veracidad del dicho popular “el mundo es un pañuelo”. Su inquietud era la siguiente: *“tomando dos personas cualesquiera en el mundo, ¿cuál es la probabilidad de que se conozcan entre sí? (...) Dados dos puntos cualesquiera elegidos al azar de este universo de 200 millones de puntos, ¿a través de cuántos puntos intermedio pasaríamos antes de que los puntos elegidos puedan ser conectados por el camino más corto posible?”* (Milgram, 2003, pp.1-17)

El primer ensayo, conocido como el estudio de Kansas, llevado a cabo en 1967 consistió en enviar una carta a una muestra de personas elegidas al azar en una ciudad de Estados Unidos, Kansas, solicitando que la remitieran a una persona objetivo, elegido también al azar y residente en un estado diferente. A los participantes se les pidió que no enviaran la carta directamente a su destinatario sino que la pasaran al amigo o conocido que ellos considerasen tuviera más posibilidades de conocerle. Cada vez que alguien remite el panfleto manda una postal a los investigadores indicando, entre otras cosas, la relación entre sí mismo y el próximo receptor. De esta forma se generaron cadenas de intermediarios que permitieron analizar algunas de las propiedades de la estructura social.

El resultado más sorprendente del experimento es que en la mayoría de las cadenas sólo fueron necesarios cinco pasos como promedio para poner en contacto a los remitentes con el destinatario¹⁸. En sus dos experimentos posteriores, el número promedio volvió a repetirse con valores cercanos: 5.2 en el Estudio de Nebraska¹⁹ y entre 5.5 y 5.9 en el Estudio de Los Ángeles²⁰. Lo asombroso de estos resultados y sus valores, entre 5 y 6, calaron en la cultura popular con la obra de teatro “Seis grados de separación” de John Guare, y el juego “A seis grados de Kevin Bacon”, consiguiendo que la teoría de “El mundo pequeño” - “*Small-World*” - de Milgram sea también conocida como la teoría de los “Seis grados de separación”.

El segundo resultado sobresaliente del estudio puso de manifiesto que prácticamente la mitad de las cadenas llegaron a su destinatario en el último paso a través de las mismas tres personas: “(...) ese *“tejido social firmemente urdido”* no parece haberse constituido por conexiones al azar, sino que existen condicionantes sociales que convierten a determinados caminos e individuos en canales preferentes de comunicación.” (Jariego, 2003, p.3)

Este hecho suponía aceptar que, además de su ubicación física, la localización ocupacional del destinatario es clave para la definición de la cadena y que, por tanto, la longitud de ésta depende más de la distancia social - podemos entender la distancia social entre dos individuos en un sistema como el número de líneas en el recorrido más corto entre uno y otro (Garnovetter, 1973, p.6)- que de la geográfica. El tercero de sus estudios, el Estudio de Los Ángeles, pretendía investigar qué ocurría cuando se modificaba la raza del destinatario, comparando una serie de mensajes enviados entre individuos de raza blanca con otras cuyo destinatario era de raza negra. Los resultados desvelaron que *“Las primeras tuvieron dos veces y media más posibilidades de completarse que las de carácter intergrupar. (...) podemos apuntar*

¹⁸ Milgram ya apuntó que había que tener en cuenta que detrás de cada uno de esos 5 intermediarios se encuentra un núcleo mucho más amplio, que cada participante a su vez tiene un pool de conocidos de 500 a 2.500 personas.

¹⁹ Travers y Milgram, 1969.

²⁰ Korte y Milgram, 1970.

que la estructura de contactos permitió confirmar la situación de segregación de la minoría negra, al tiempo que se identificaron puentes –vinculados a los roles de determinados individuos- que hacían posible la comunicación «inter-racial».” (Jariego, 2003, p.4)

La naturaleza de esos puentes fue analizada por Granovetter en su artículo sobre “La fuerza de los lazos débiles”, gracias a la información suministrada por los participantes en la cadena, que tenían que clasificar sus contactos como “amigo” –vínculo fuerte- o “conocido” –vínculo débil- y fijándose en lo que para él era un punto crucial, el primer envío de un blanco a un negro. Concluyó que cuándo se marcaban como “conocido” la cadena se completaba en el 50% de los casos y que este porcentaje caía hasta la cifra del 26% cuando se señalaban como “amigos”. Estas informaciones corroboraban su teoría: *“Así los lazos interraciales más débiles pueden verse como más efectivos en distancias sociales mediante puente.”* (Granovetter, 1973, p.9)

Posteriormente, en un estudio de de S Peter Sheridan Dodds, Roby Muhamad y Duncan J. Watts, se intentó replicar el ensayo de Milgram para poner a prueba la validez de sus conclusiones²¹. En el resultado, aunque el experimento se desarrollaba mediante contacto por correo electrónico entre los más de 60.000 implicados repartidos en 166 países, se llegó a la conclusión de que cualquier individuo elegido al azar puede alcanzar a su destinatario en pocos pasos: *“Finalmente, se llega a una estimación de entre 5 y 7 pasos, dependiendo de la separación geográfica entre la fuente y el objetivo, y, a diferencia de las conclusiones de Milgram, no hay evidencias de que los mensajes se canalicen a través de individuos socialmente prominentes. La búsqueda en redes sociales globales aparece como un fenómeno bastante «igualitario».*” (Jariego, 2003, p.5)

²¹ DODDS, Peter Sheridan, MUHAMAD, Roby y WATTS, Duncan J. (2003). “An experimental study of search in global social networks”. *Science*, 301, 827-829. Mencionado en Jariego (2003).

La importancia de la fuerza de los vínculos débiles junto con la facilidad de búsqueda en la red que propician los medios de comunicación electrónicos son atributos constantes y de gran peso en la estructura de las redes que se forman en Internet, afectando por tanto estas características, a la forma de la difusión de información y/o conductas en este canal.

2.3.3. Tres grados de influencia

¿Tenemos capacidad de influencia sobre todas las personas que se encuentran a determinada distancia social de nosotros? Las investigaciones de N.A. Christakis y J. H. Fowler han demostrado que el grado de influencia en las redes sociales obedece a lo que ellos denominan “Regla de los tres grados de influencia”: *Todo lo que hacemos o decimos tiende a difundirse –como las olas- por nuestra red y tiene cierto impacto en nuestros amigos (un grado), en los amigos de nuestros amigos (dos grados) e incluso en los amigos de los amigos de nuestros amigos (tres grados). Nuestra influencia se disipa gradualmente y deja de tener un efecto perceptible en las personas que se encuentran más allá de tres grados de separación.*” (Christakis y Fowler, 2010, p.41)

Los autores argumentan que la influencia que podemos ejercer en los demás – y que los demás ejercen sobre nosotros- termina por agotarse y se razona con las siguientes evidencias: (Christakis y Fowler, 2010, pp.41-43)

- Explicación de la decadencia intrínseca. La fidelidad de la información va decreciendo a medida que se transmite.
- Explicación de la inestabilidad de la red. Los vínculos en la red no duran siempre y son aquellos con los que no mantenemos una conexión directa, los menos estables.
- Explicación a efectos evolutivos. En nuestro pasado homínido, en el que evolucionamos en grupos pequeños, nadie estaba a cuatro grados de separación de nosotros.

Debido a la estructura de las redes sociales naturales, cada uno de nosotros está conectado con miles de personas y el alcance de lo que podemos causar en otros es de un volumen numérico extraordinario. Por ejemplo, con sólo veinte contactos de media podré estar a tres grados de separación de una población cercana a las ocho mil personas, dado que esos veinte contactos directos (primer grado) a su vez tendrán otros veinte contactos directos (segundo grado), los cuáles a su vez tendrán veinte contactos (tercer grado): $20 \times 20 \times 20 = 8.000$.

La pregunta clave es si estos tres grados de influencia se mantienen en las redes sociales en Internet, como Facebook o Twitter, o si se podrían superar debido a que en ellas las “explicaciones” arriba citadas no se cumplen como tal:

- La fidelidad de la información se mantiene por su carácter electrónico (las palabras y las imágenes pueden transmitirse siempre fieles a su forma original).
- Los vínculos tenderían a ser más estables al no quedar sujetos a las condiciones particulares espaciales o temporales de sus nodos (las personas permanecen vinculadas y en contacto aunque alguna de ellas traslade su lugar de residencia).

Si los experimentos de Milgram revelaron que todos estamos conectados, la regla de “Los tres grados de influencia” incide en el potencial de contagio de todo lo que fluye o circula por la red²². Por su importancia en la difusión o propagación de ideas y conductas en las redes sociales, dedicaremos un epígrafe al conocimiento y reflexión sobre el contagio.

2.3.4. La fuerza de los vínculos débiles

²² Relacionadas con el concepto contagio encontramos la teoría de “Memes” (DAWKINS, Richard 2006. *The selfish gen*. Oxford University Press) y la teoría de “La difusión de innovaciones” (ROGERS, Everet M. (1962) *Diffusion of innovations*. Free Press of Glencoe).

El sociólogo estadounidense Mark Granovetter destacó en los años setenta por sus modernas teorías en el campo de la sociología, concretamente en la que anuncia el título de este epígrafe, en su título original *“The Strength of Weak Ties”* (1973), y *“Tipping Points/threshold models”*, ensayo en el que experimentó sobre las conductas del ser humano en la propagación de las modas²³.

Partimos del acuerdo de que un vínculo dado puede ser fuerte (amigo), débil (conocido) o ausente (incluye falta de cualquier relación y lazos sin significado substancial, como por ejemplo, una cordial relación entre vecinos). Granovetter define la fuerza de un vínculo como *“una (probablemente lineal) combinación del tiempo, la intensidad emocional, intimidad (confianza mutua) y los servicios recíprocos que caracterizan a dicho vínculo”*. (Granovetter, 1973, p.2)

La primera consecuencia de la fuerza de los vínculos es que en un trío (A, B y C), en el que A está fuertemente unido con B y con C, existen muchas posibilidades de que B y C sean parecidos a A, y dada dicha similitud y el hecho de que A pasará mucho tiempo con B y con C y que por tanto compartirán algunos momentos, B y C es muy probable que interactúen. Esta probabilidad es menor cuando los lazos de A con ellos son de fuerza débil. Sin embargo, en este caso B y C querrán que A siempre sea congruente con cada uno de ellos, independientemente, creándose una negatividad psicológica, caso que no ocurre en las relaciones de vínculos débiles.

Sobre las implicaciones y consecuencias de las relaciones de vínculos débiles analizadas por Granovetter, la que más nos interesa en la dedicada al papel de estos en la difusión de la información. En ella incluye el término “puente”, una línea en un sistema que proporciona el único camino entre dos puntos: *“En general, cuando cada persona tiene muchos contactos, un puente entre A y B proporciona la única ruta”*²⁴ a

²³ Este modelo de umbral de la conducta social se propuso anteriormente por Thomas Schelling y se popularizó en 2002 Malcolm Gladwell con su libro *“The Tipping Point: How Little Things Can Make a Big Difference”*, del que hablaremos más adelante.

²⁴ El mismo autor aclara que en grandes redes, es raro que un vínculo específico proporcione el único camino entre dos puntos. Puede aplicarse a las redes en Internet.

los largo de la cual la información o influencia puede ir desde cualquier contacto de A a cualquier contacto de B, y, consecuentemente, desde cualquiera conectado indirectamente a A hasta cualquiera conectado indirectamente a B²⁵.” (Granovetter 1973, p.4)

El autor argumenta que cualquier cosa que pueda ser difundida podrá, como consecuencia de lo expuesto en el párrafo anterior, llegar a un gran número de personas y atravesar una gran distancia social cuando se experimentan vínculos débiles antes que fuertes: *“Si alguien cuenta un rumor a todos sus amigos cercanos, e igualmente hacen ellos, muchos oirán el rumor una segunda y tercera vez, ya que aquellos que están unidos por vínculos fuertes tienden a compartir amigos. Si la motivación para extender el rumor se debilita cada vez que éste se cuenta, entonces el rumor que se mueva a través de vínculos fuertes estará probablemente limitado a unos pocos cliques o pandillas que si fuera vía vínculos débiles; los puentes no serán cruzados.”* (Granovetter, 1973, p.6). Su conclusión final es que los vínculos débiles son indispensables para las oportunidades individuales y para su integración en la comunidad y que los vínculos fuertes llevan a una fragmentación total.

Así como ha quedado probado por algunos estudios que la centralidad es un atributo relevante de un nodo a la hora de adoptar un conducta determinada y difundida por él, los vínculos fuertes y más concretamente en el caso de Internet, no han demostrado su preferencia como atributo en estos casos. Si bien se consideraba que los vínculos fuertes eran más confiables y creíbles en el momento de adoptar un nuevo comportamiento, en este nuevo canal los resultados no lo avalan. El ensayo “Electronic Word-of-Mouth in Online Environments: Exploring Referral Network Structure and Adoption Behavior” demostró que en Internet los pesos de vínculos fuertes y de los vínculos débiles se igualan en valor: *“This was an interesting result,*

²⁵ Esta afirmación coincide con la teoría de “Los tres grados de influencia” expuesta por Christakis y Fowler (2010) *Conectados, el sorprendente poder de las redes sociales y cómo nos afectan*. Editorial Taurus, y sobre la que incidiremos más adelante.

implicating that all connections in virtual electronic environments are equal in their effectiveness and persuasiveness.” (Vilpponen et al., 2006, p.73)

Los politólogos Dan Carpenter, Kevin Esterling y David Lazer estudiaron las redes sociales de los lobbies en algunas áreas, como la energía y la sanidad²⁶. Descubrieron que si bien para ellos son importantes los lazos fuertes entre sus homólogos ideológicos, sus logros dependen también del modo en que se comporte la totalidad de la red en la que se mueven: *“Esto quiere decir que los lobbistas que más éxito tienen en sus empresas son aquellos con vínculos más débiles, a saber, quienes tienen más amigos con más amigos en los vínculos de poder. Los vínculos fuertes ayudan pero los débiles son más útiles porque aumentan el número total de conexiones.*” (Christakis y Fowler, 2010, p.213)

Se ha demostrado la importancia de los vínculos débiles por lo que podemos concluir este epígrafe afirmando que si aceptamos la realidad de que Internet es un medio/canal en el que las redes sociales que se establecen favorecen y aumentan el número de vínculos débiles entre sus actores, se convierte en el medio ideal para la difusión de información y la adopción de conductas similares y tempranas.

2.3.5. El contagio en la red: memes y difusión cultural

El concepto de contagio se ha extendido desde su origen en el ámbito de la salud dedicado al estudio de la difusión de epidemias hasta su uso en una amplia variedad de procesos que se propagan en las redes, con una fuerte carga de tipo social, como la moda, las opiniones políticas y la adopción de nuevas tecnologías.

Uno de los primeros contagios con que nos encontramos es el “contagio emocional”²⁷. Cómo nos sentimos depende de cómo se sienten las personas más

²⁶ CARPENTER, Dan, ESTERLING, Kevin y LAZER, David. “Friends, brokers and Transitivity: Who Informs Whom in Washington Politics” *Journal of Politics* 66 (2004), pp. 224-246. Mencionado por Christakis y Fowler (2010).

²⁷ Véase como referencia el informe de Rankin y Philip sobre la “epidemia de risa” que brotó en 1962 en un internado femenino de Bukoba-Tanzania. RANKIN, A. M. y PHILIP, P. J. (1963). “An

cercanas a nosotros. Si nuestros amigos son felices, nuestras probabilidades de felicidad aumentan, y viceversa. Algunos autores consideran este hecho fruto de la evolución del ser humano: *“Es probable que las emociones y el contagio emocional surgieran primero para facilitar el vínculo madre-hijo y que luego se extendieran a los miembros del clan y, en última instancia a quienes no pertenecieran a éste (...). Adoptar rápidamente el estado emocional de los demás puede incrementar las posibilidades de supervivencia. De hecho, se cree que las emociones positivas pueden funcionar especialmente bien para aumentar la cohesión del grupo («Estoy contento», «Quédate conmigo») y que las emociones negativas pueden funcionar bien como herramienta de comunicación («Huelo a humo», «Tengo miedo»).”* (Chistakis, 2010, p.50)

Siguiendo teorías biológicas evolutivas, Richard Dawkins introdujo en 1975 el término “meme” al que definía como la unidad de información cultural que puede ser transmitida de una mente a otra, de una persona a otra, de una generación a otra; un modelo o patrón de información retenido en una memoria que es capaz de ser copiado en otra memoria individual²⁸. La voz “meme” puede aplicarse a todo aquello que pueda ser retenido por un sujeto como ideas, conocimiento, creencias, costumbres, imágenes, etc.

Junto al término “meme”, Dawkins también acuñó la palabra “memética” (del inglés “memetic”) o ciencia que estudia la replicación, propagación y evolución de los “memes”²⁹. Cuando un “meme” se convierte en éxito es porque ha sido capaz de replicarse en muchas personas, propagándose entre nosotros durante un cierto espacio de tiempo.

Epidemic of Lauching in the Bukoba District og Tanganyka”. *Central Africa Journal of Medicine* 9, pp. 167-170.

²⁸ DAWKINS, Richard. (2006). *The selfish gen*. Oxford University Press.

Etólogo, zoólogo, teórico evolutivo y escritor de divulgación científica británico (de origen keniano) que ocupa la «cátedra Charles Simonyi de Difusión de la Ciencia» en la Universidad de Oxford.

²⁹ Ciencia descrita y desarrollada en *Journal of Memetics: Evolutionary Models of Information Transmission*. El término “memetic” enlaza con “mimetic”, en español “mimético” o que imita por mimetismo.

¿Qué hace que un “meme” obtenga dicho éxito? Francis Heylighen describe los estadios por los que tiene que sucederse un “meme” para lograrlo: (Heylighen, 1999, pp.3-6)

- Asimilación por un individuo. Debe ser notado, comprendido y aceptado por el anfitrión o host. Supone que es lo suficientemente sobresaliente para atraer la atención del huésped y que el anfitrión reconoce al “meme” como algo que se puede representar en su sistema cognitivo.
- Retención en la memoria. Debe permanecer algún tiempo en la memoria inicial. Cuanto más tiempo permanece en ella más oportunidades tendrá para difundirse en mayor medida al infectar otras memorias o nuevos hosts. La retención dependerá de la importancia de la idea para el individuo, y cuantas veces ésta se repite, ya sea por percepciones recurrentes o por ensayos internos.
- Expresión comprensible de este individuo para otros. Expresión en el lenguaje, comportamiento u otra forma que pueda ser percibido por los demás. La manera más evidente de expresión es el lenguaje. Otros medios comunes para la expresión meme son texto, imágenes, y el comportamiento. No requiere la decisión consciente de expresarlo (puede ser el modo en que utilizamos algo).
- Transmisión. Para llegar a otro individuo necesita un soporte físico o medio suficientemente estable para transmitir la idea con la mínima pérdida o de información (lenguaje, libros, CD's, etc.)

Si bien en el estudio de los “memes” no se aplican teorías de redes en su transmisión, su introducción en este trabajo es interesante puesto que queda implícita la idea de contagio. Y en las redes el contagio es un factor del mayor interés.

Relacionado con el concepto de “meme” encontramos diversos estudios aplicados a la comprensión de los factores que intervienen en la “difusión cultural” o difusión y contagio de comportamientos. Ya vimos en el epígrafe dedicado a “La

fuerza de los vínculos débiles” como Granovetter también había estudiado las conductas del ser humano en la propagación de las modas y cómo Malcom Gladwell popularizó estas teorías en 2002 con su libro *“The Tipping Point: How Little Things Can Make a Big Difference”*³⁰. En él, Gladwell afirma que las ideas, los productos, los mensajes y las conductas se extienden entre nosotros igual que los virus³¹. Las conclusiones que nos acerca y más nos interesan son las siguientes:

- Las conductas son contagiosas.
- La ley de pocos. Partiendo de unos pocos la información puede transmitirse de forma efectiva a una escala mucho mayor.
- Pequeños gestos. Pequeños cambios producen grandes efectos.
- Tres tipos de nodos en la red. Conectores (personas especiales capaces de conectarnos con cualquiera), vendedores (transmiten con entusiasmo sus propias emociones al resto) y “mavens” (acumuladores de conocimiento que luego ofrecen a los demás)

Sin duda una teoría de gran relevancia en el campo del análisis de la difusión cultural es la denominada teoría de la “Difusión de innovaciones”, que intenta descubrir cómo y a qué velocidad se difunden las nuevas ideas en un entorno social. Fue popularizada por Everett M. Rogers en 1962 con su escrito *“Diffusion of innovations”* en el que identifica y analiza los cuatro elementos principales que actúan en una difusión: (Rogers, 1962, pp.35-37)

- Innovación. Una innovación es una idea, práctica u objeto percibido como algo nuevo por un individuo u otra unidad de adopción. De sus características depende su grado de adopción.
- Canales de comunicación. Los medios de comunicación son efectivos creando conocimiento sobre una innovación mientras que

³⁰ En el lenguaje de los epidemiólogos “The Tipping Point” es el momento en que un fenómeno estable de bajo grado o grado controlado se convierte en una crisis de salud pública. Una epidemia.

³¹ Ejemplo: el éxito obtenido por un modelo de zapatos *Hush Puppies* que ya había dejado de fabricarse.

los individuos son efectivos formando y modificando una idea, y por tanto, influenciando sobre su adopción o rechazo por otros.

- Tiempo. Tiempo de decisión, tiempo de capacidad de innovación y tiempo de adopción. Distingue cinco categorías de adoptadores: innovadores, adoptadores tempranos, mayoría temprana, mayoría tardía, y rezagados.
- Sistema social. La estructura social y de comunicación de un sistema facilita o impide la difusión de la innovación.

Rogers afirma que los adoptadores tempranos son marginales³² y hasta se les percibe como desviados –hecho por el que no serían muy imitados por otros– y el grupo de los adoptadores más tardíos están más integrados en el sistema social. Granovetter dedica parte de su estudio sobre los vínculos débiles a la difusión, interesándose por el papel de los centrales y marginales en el proceso de adopción de innovaciones: *“Becker (1970) intenta resolver la pregunta sobre si los innovadores tempranos son “centrales” o “marginales” refiriéndose a los “riesgos percibidos de adopción de una innovación dada”. Su estudio sobre las innovaciones de salud pública demuestra que cuando un nuevo programa es considerado como seguro y sin controversias (como la droga de Coleman et al.), las figuras centrales son las primeras en adoptarlo; si no, la marginales lo hacen (p. 273). Explica la diferencia en términos de un deseo mayor de las figuras “centrales” de proteger su reputación profesional.”* (Granovetter, 1973, p.7)

Completando las conclusiones de Rogers, Granovetter afirma que los vínculos débiles, una vez más, juegan un papel crucial en la difusión de innovaciones. Así, a mayor número de vínculos débiles, mayor éxito para difundir una innovación difícil: *“Debemos pensar que, ya que hay resistencia mayor a una actividad arriesgada o desviante que a una normal y segura, un mayor número de personas tendrá que exponerse a ella y adoptarla en los primeros pasos antes de que*

³² Centrales son aquellos nodos que, en un estudio, reciben muchas elecciones y marginales aquellos que reciben menos.

se extienda en una reacción en cadena. Los individuos con muchos vínculos débiles, según mis argumentos, están mejor situados para difundir una innovación tan difícil, ya que algunos de estos vínculos son puentes locales.” (Granovetter, 1973, p.8)

En contraste con el modelo de un solo paso (*one-step-flow*), o “hipodérmico”, modelo en el que los individuos son tratados como objetos atomizados por la influencia de los medios de comunicación (Bineham, 1988)³³, el modelo de formación de la opinión pública formulado por Katz y Lazarsfeld (1955) supuso una revolución en el campo de la sociología, siendo que durante décadas la mayoría de los estudios dedicados al análisis de la influencia, reafirmaban su teoría en los procesos de difusión y estudios de mercado. Su modelo pudo suponer un “lavado” de cara para los medios de comunicación de masas, al afirmar que en un variedad de escenarios de toma de decisiones, que van desde la política al personal, los individuos pueden verse más influenciados por la exposición entre sí que a los medios. Según su teoría, una pequeña minoría de “líderes de opinión” actúan como intermediarios entre los medios de comunicación y la mayoría de la sociedad. Dado que la información, y por tanto la influencia, fluye desde los medios de comunicación a través de los líderes de opinión a sus respectivos seguidores, Katz y Lazarsfeld (1955) llamaron a su modelo de comunicación “*two-step-flow*”.

Más de cincuenta años después, Watts y Dodds (2007) publicaron los resultados de su estudio sobre esta “hipótesis de la influencia” definida por Katz y Lazarsfeld (1955), utilizando una serie de simulaciones por ordenador de los procesos de influencia interpersonal para determinar si los “influyentes” o bien conectados son los nodos más propensos a desencadenar una difusión en cascada. Sus resultados sugieren que los nodos definidos como influyentes no son más apropiados para conseguirlo que la media de los nodos, contradiciendo las teorías más populares que durante décadas así lo afirmaban. En los modelos estudiados, de

³³ BINEHAM, J. L. (1988), “A Historical Account of the Hypodermic Model in Mass Communication,” *Communication Monographs*, 55, 230–46. Citado por Katz y Lazarsfeld (1995).

hecho, la mayoría de los cambios sociales no están impulsados por influyentes sino por individuos fácilmente influenciables que, a su vez, influyen en otras personas fácilmente influenciables. En base a estos resultados, argumentan que, a pesar de que sus modelos son los mejores en una representación simplificada y parcial de una realidad compleja, sirven para destacar que las reclamaciones sobre el papel relevante de los influyentes deben basarse en supuestos cuidadosamente especificados sobre quién influye en quién y cómo. Además, demostraron que la probabilidad de adoptar una innovación se incrementa cuando un grupo de usuarios, y no sólo uno, repite el mismo mensaje.

“Under most conditions that we consider, we find that large cascades of influence are driven not by influentials but by a critical mass of easily influenced individuals. Although our results do not exclude the possibility that influentials can be important, they suggest that the influentials hypothesis requires more careful specification and testing than it has received.”
(Watts y Dodds, 2007, p.441)

Ilustración 1. Visualización de la hipótesis de la influencia de Watts y Doods (2007)

FIGURE 1
SCHEMATIC OF THE TWO-STEP FLOW MODEL
OF INFLUENCE

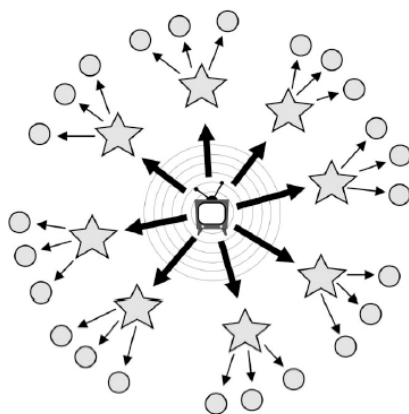
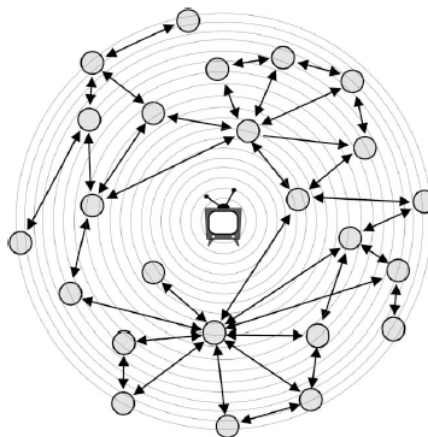


FIGURE 2
SCHEMATIC OF NETWORK MODEL OF INFLUENCE



El proceso de difusión que sigue un fenómeno que llega a convertirse en “viral” es cualitativamente diferente del que se sucede en aquellas situaciones en las que el mensaje o el producto se desvanecen rápidamente en la oscuridad o no llegan a alcanzar la dominancia. Las dinámicas de la difusión social dependen estrechamente de los escasos vínculos (posibilitan las interacciones) existentes entre grupos con individuos densamente conectados (comunidades).

Una de las hipótesis que Colbaough y Glass (2009) plantearon es que la intensidad de las interacciones entre comunidades es un indicador de la magnitud y el alcance de una difusión social³⁴. Vigilando la actividad que llevaron a cabo los activistas del partido socialista sueco SPD, viajando a lo largo del país para conseguir seguidores, la primera conclusión que alcanzaron fue que como consecuencia de dicha actividad se formó una red de conexiones que vinculaban las hasta ahora dispersas comunidades. Y más sorprendente, que las interacciones que se establecían entre comunidades (inter-community) suponían una mejor predicción de los miembros del SPD para el año siguiente en un distrito que el número de miembros actuales: *“Thus inter-community interaction is more predictive than either process intrinsics or standard diffusion measures (which focus on intra-community dynamics).”* (Colbaough y Glass, 2009, p.6)

Dicha hipótesis, la afirmación de que los vínculos entre comunidades son determinantes en el proceso de difusión, fue refutada por los mismos autores en la obtención de un modelo de predicción para la difusión de la epidemia del SARS a nivel internacional durante los años 2002 y 2003, tomando como referencia los datos de viajeros de vuelos comerciales entre países³⁵. A pesar de su simplicidad, la

³⁴ Debido a sus características y el método de ensayo, la segunda hipótesis se incluye en el apartado dedicado a Internet

³⁵ SARS, Siglas para denominar al “Síndrome respiratorio agudo severo”, neumonía atípica que apareció por primera vez en noviembre de 2002 en la provincia de Guangdong, China. Se propagó a las vecinas Hong Kong y Vietnam a finales de febrero de 2003, y luego a otros países a través de viajes por medio aéreo o terrestre de personas infectadas.

Datos que se obtuvieron en la página Web de IATA, <http://www.iata.org/index.htm>

simulación de su modelo mostraba un alto grado de coincidencia con la propagación del SARS. Las relaciones entre comunidades dispares son clave para la difusión:

“(...) because inter-community interaction is a key element of social network diffusion, this dynamics is modeled more carefully than intra-community interaction.” (Colbaough y Glass, 2009, p.7)

Como resultado de este epígrafe dedicado al contagio y a la difusión cultural, se desprende que si el contagio es posible, la popularidad no siempre es sinónimo de superior influencia y los lazos débiles son una pieza clave en el proceso, las redes sociales configuradas en los servicios, que a tal efecto, ofrece Internet, contienen los ingredientes óptimos para mejorar un proceso de difusión: alto número de conexiones entre individuos muy diversos y proliferación de vínculos débiles.

2.4. El comportamiento de Internet como red

2.4.1. Características de Internet según teorías de redes

Podemos considerar el informe de Dodds, Muhamad y Watts, “An experimental study of search in global social networks” publicado en 2003 en *Science*, como el primer análisis de grandes redes llevado a cabo en Internet. Una de las conclusiones más relevantes del estudio es que *“no hay evidencias de que los mensajes se canalicen a través de individuos socialmente prominentes. La búsqueda en redes sociales globales aparece como un fenómeno bastante “igualitario” (pág. 828), en lo que se refiere a la distribución de los mensajes. Además, la presencia de individuos altamente conectados es de escasa relevancia para las estrategias de búsqueda que se despliegan durante el experimento.”* (Jariego, 2003, p.6) Según los autores, esta particularidad se debe a que la estructura de la red por sí misma no lo es todo y que en los análisis hay que tener en cuenta la vertiente psicológica de los participantes, quienes tendrán en cuenta no sólo la centralidad de un nodo sino su facilidad de acceso o localización, y, a su vez, la posibilidad o conexión de éste con

el objetivo final. Se desprende que aunque los amigos son el tipo de lazo al que se recurre en primer lugar, en las cadenas completas -las que alcanzan a su destinatario- sobresalen los vínculos con conocidos por motivos profesionales y las relaciones casuales o no muy cercanas, aportación última que corrobora la importancia de los lazos débiles en la conectividad de las redes sociales.

En su libro, “Linked, how everything is connected to everything and what it mean”, Alfred-László Barabási (2003)³⁶ argumenta que Internet, contrariamente a lo que se pensaba, no es una red aleatoria³⁷ en la que los vínculos se establecen un tanto al azar, sino que su estructura corresponde a una “red de libre escala” en la que unos pocos individuos reciben miles o millones de vínculos, nodos a los que el autor denomina *hubs* y a los que considera cruciales para el mantenimiento de la red por servir de conectores entre grupos de usuarios. Internet, dice, es una red con pocos *hubs* que obedece a la ley de potencias³⁸: a medida que crece el número de links disminuye el número de nodos: *“La aportación de Barabási y Albert (1999) consistió precisamente en poner el acento en el crecimiento de la red, de forma que un mecanismo de “enlace preferencial” lleva a que los nuevos vínculos se asocien preferentemente con aquellos que ya están particularmente bien conectados. De esa forma, un pequeño número de ejes -esto es, nodos que destacan por su elevado número de conexiones- son fundamentalmente responsables de la topología de la*

³⁶ Apoyándose en el ensayo realizado junto con Reka Albert en 1999, “Emergence of scaling in random networks” publicado en *Science*.

³⁷ Modelo propuesto por los matemáticos húngaros Paul Erdős y Alfréd Rényi en el que los nodos se conectaban primero al azar y más tarde de par en par.

³⁸ En los años "30, el lingüista George Kingsley Zipf descubrió una tendencia poco usual en el lenguaje: en cualquier obra literaria: la segunda palabra más usada, aparece la mitad de veces en el texto que la que se usa más. La tercera más usada figura tres veces menos, y así sucesivamente. Zipf descubrió lo que hoy día se conoce ampliamente como la ley de potencias: Una ley de potencias es un tipo especial de relación matemática entre dos cantidades. Aplicado a la estadística, si estas dos cantidades son la variable aleatoria y su frecuencia, en una distribución de ley de potencias, las frecuencias decrecen según un exponente cuando la variable aleatoria aumenta. <http://redactores.zoomblog.com/archivo/2012/02/24/ley-de-potencias.html>. (Consultado el 21/02/2012)

red: ellos determinan la estabilidad de la estructura, su comportamiento dinámico y su grado de resistencia (Barabási, 2002). ” (Jariego, 2003, p.7)

Albert y Barabási se enfrentaron con una limitación que ya había sido manifestada por otros investigadores o rastreadores Web y que consiste en que los robots no eran capaces de indexar todas las páginas o información que circula por Internet: *“Por más impresionante que sea la capacidad de indexación de los motores, la web crece infinitamente más rápido que su posibilidad de analizarla. Y para peor -además de la web superficial a la que alcanzan estos robots- tenemos una web profunda que es por lo menos 550 veces más grande que la anterior, con lo que la asimetría entre existente y visible se ahonda muchísimo más.” (Rodríguez, 2004)*

¿Partiendo de una página concreta, hasta dónde nos llevarán sus vínculos? La indexación de los robots y su posterior análisis ha puesto de manifiesto que los links (enlaces a nuevas URL's) son asimétricos, unidireccionales, y por ello aquellos nodos que reciben muchos links son “vistos” y los que no, pueden quedar invisibles³⁹: *“Lo que estas investigaciones dejaron en claro es que para recorrer la web no sólo cuentan los incentivos económicos sino que por su propia topología la red tiene una estructura que tanto enfatiza su visibilidad como su invisibilidad. La W3 es una red libre-de escala (del mismo tipo como lo es Internet en su conjunto, o del tipo el ganador se queda con todo que es tan propio en la economía de las estrellas).” (Rodríguez, 2004)*

Las conclusiones de Barabási se remitían al estudio de Internet (blogs, medios, instituciones, etc.). Posteriormente, en la conferencia “Redes y propiedades emergentes”⁴⁰, apuntó que las comunidades online corroboran esta visión de la red en

³⁹ En la captura de *tweets* para su investigación, Romero et al. (2011) denotaron que los 22 millones de *tweets* acumulados representaba un 1/15 de la actividad completa de Twitter en ese momento concreto.

⁴⁰ Dictada en el ciclo “Los debates abiertos de la Fundación Telefónica” el 15 de septiembre de 2009.

la que algunos usuarios tienen pocos amigos y otros pueden llegar a tener miles. Estos usuarios se convertirían, por tanto, en *hubs* de la comunidad online.⁴¹

En el año 2000, Cass R. Sunstein⁴², realizó una investigación similar eligiendo 60 sitios políticos con la que alcanzó conclusiones que refuerzan la direccionalidad de los vínculos, la puesta en marcha de mecanismos de autorrefuerzo y la creación de micro-comunidades en la red: “*apenas 15% remitían a sitios con opiniones contrarias. Al revés un 60% mantenía links hacia sitios afines. Y lo mismo pasó con otro estudio dedicado a analizar el discurso democrático en la red.*” (Rodríguez, 2004)

En cierta contraposición con la mencionada “Ley de Potencias” de Barabási, en los resultados del ensayo “*Electronic Word-of-Mouth in Online Environments: Exploring Referral Network Structure and Adoption Behavior*” de Antti Vilpponen, Susanna Winter, and Sanna Sundqvist (citado anteriormente), se concluye que aunque la centralidad y los lazos fuertes son determinantes en la transmisión y adopción de una innovación, la difusión boca a boca en una red electrónica se revela como bastante descentralizada⁴³: “*strong ties were more numerous and more*

⁴¹ La Ley de Potencias, que puede denotar un cierto carácter negativo al asumir que unos pocos alcanzan popularidad y por tanto capacidad de influencia, puede aplicarse a los nuevos mercados, adquiriendo un significado positivo. Se trata del fenómeno *Long Tail* que tiene su origen en el artículo de Chris Anderson, “The Long Tail” publicado en Octubre del 2004 en Wired. La Ley de Potencias se ve disminuida en tanto que favorece la aparición de nuevos mercados: existe un mercado centrado en el alto rendimiento de pocos productos, el denominado mercado de masas; y por otro, un mercado innovador y cada vez más influyente basado en la suma o acumulación de todas las pequeñas ventas de muchos productos, que puede igualar o superar al primero; éste representan a un nuevo modelo de nicho de mercados, representado por la cola de la conocida gráfica de distribución estadística. La teoría de las “economías de larga cola” (*Long Tail Economies*) expone que la cultura y la economía se alejan de los mercados y productos que encabezan la curva de la demanda (los más populares) y se expanden en la parte alargada de la curva (los minoritarios). Constituyen el antiguo mercado de masas y el nuevo nicho de mercados. Aplicado a los usuarios influyentes en Internet, junto con una minoría influyente conviviría una numerosa masa de usuarios que propagarían sus opiniones entre sus “clientes” o usuarios conectados.

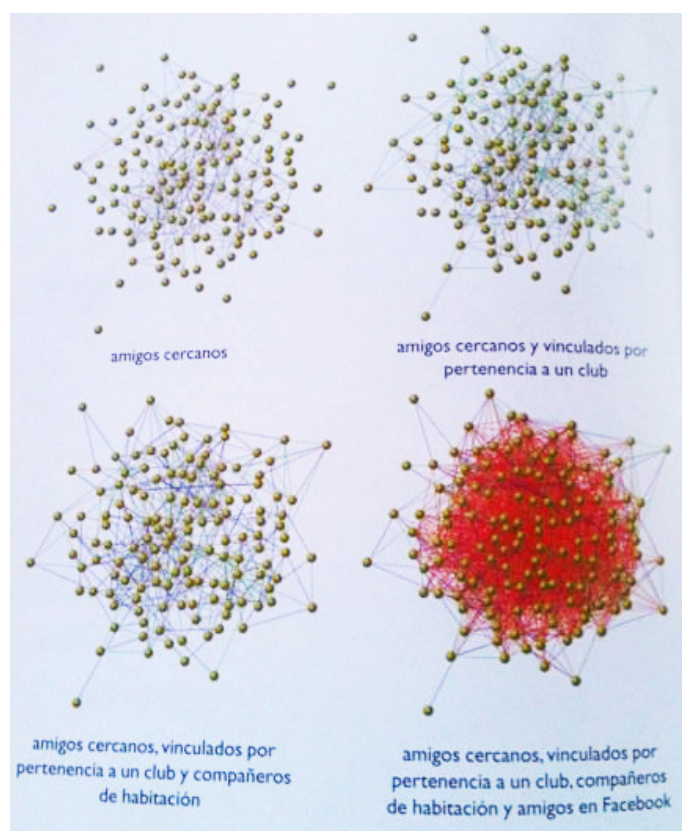
⁴² Autor del libro *República.com* (2003). Editorial Paidós. Y del libro *Rumorología* (2010). Editorial Debate.

⁴³ Sin embargo, el estudio de influencia en Twitter de Cha et al. (2010) confirma la ley de potencias, indicando que los *Top Twitter User* poseen una cantidad desproporcionada de la influencia global en esa red.

influential as information sources but were less likely to be actively sought out,” (Vilpponen et al., 2006, p.68), lo cual puede estar relacionado con la diferencia existente entre esta comunicación electrónica y la comunicación cara a cara: *“Empirical studies have found that computer-mediated communication is more task-oriented, less emotional, and less personal than face-to-face communication (Hiltz, Johnson, and Turoff 1986). Actually, in virtual communities the communicating consumers may not know each other, as the use of pseudonyms is a common habit. Thus, it can be assumed that word-of-mouth behavior in electronic virtual environments is somewhat different from traditional face-to-face word-of-mouth.”* (Vilpponen et al., 2006, p.68)

En una representación gráfica de las diferentes conexiones que se establecen entre una red de amigos cercanos, la misma red con muestras de vínculos por pertenencia a un club, las condiciones anteriores más el dato compañeros de habitación y, por último, todas las coincidencias anteriores a la que se suma la particularidad de ser amigos en Facebook, Christakis y Fowler, comprobaron que *“las conexiones del mundo virtual superan con creces las conexiones en persona, e incluso llegan a oscurecer las relaciones reales.”* (Christakis y Fowler, 2010, pp.216-217)

Ilustración 2. Visualizaciones de diferentes redes realizadas por Christakis y Fowler, (2010)



Una de las hipótesis que Colbaough y Glass (2009) plantearon es que la dispersión de un fenómeno de difusión entre diferentes comunidades es un fiel indicador temprano de que la propagación va a ser significativa. La publicación en el diario danés *Jyllands-Posten* en septiembre de 2005 de caricaturas que ridiculizaban o se burlaban de Mahoma, junto la controversia que levantaron los comentarios del Papa Benedicto XVI en septiembre de 2006 y la reedición de dichas caricaturas en diferentes diarios en febrero de 2008, desencadenaron una serie de protestas. Los autores modelaron un gráfico de la “blogosfera” (vínculos entre los propios blogs y vínculos entre los conceptos que aparecen en sus *post*⁴⁴) para estudiar el modo en que

⁴⁴ Cada una de las unidades de publicación que el autor va introduciendo secuencialmente en su blog.

se dichas protestas se fueron sucediendo y magnificando⁴⁵. Se halló que en aquellos sucesos en los que la difusión en los blogs fue temprana, resultaron ser los eventos que mayor movilización social desarrollaron:

“(...) early diffusion of mobilization-related activity (here blog discussions) across disparate social communities can be a useful early signature of successful protest events.” (Colbaough y Glass, 2009, p.9)

Ilustración 3. Ejemplos de las movilizaciones blog para “Pope case” y “Danish cartoons” de Colbaugh y Glass (2009)

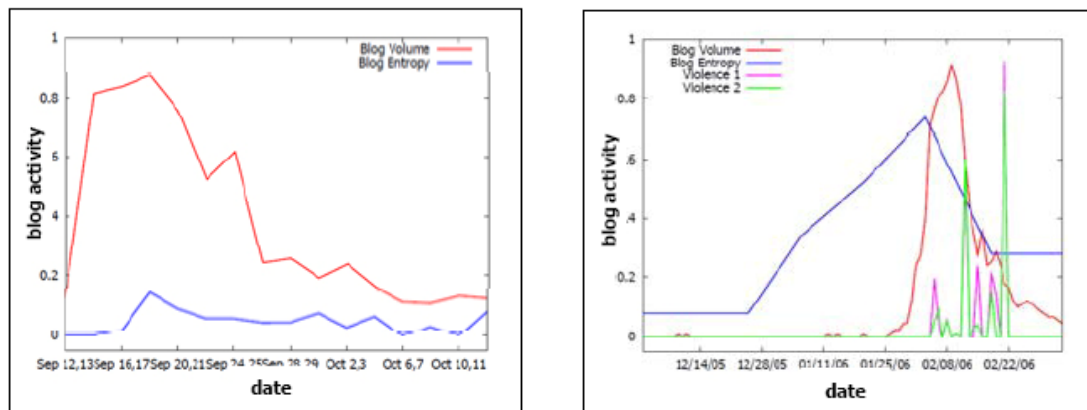


Fig. 4. Sample results for mobilization/protest case study. The illustrative time series plots shown correspond to the pope event (left) and first Danish cartoons event (right). In each plot, the red curve is blog volume and the blue curve is blog entropy; the Danish cartoon plot also shows two measures of violence (cyan and magenta curves). Note that while the volume and violence data are scaled to allow multiple data sets to be graphed on each plot, the scale for entropy is consistent across plots to enable cross-event comparison.

⁴⁵ Para profundizar en la forma de difusión en la blogosfera, pueden consultarse los siguientes estudios:

- GRUHL et al. (2004). “Information Diffusion Through Blogspace”. *Proceedings of the International World Wide Web Conference* (basado en el contenido de los posts de los blogs).
- ADAR, Eytan y ADAMIC, Lada A. (2004). “Information Dynamics LabTracking Information Epidemics in Blogspace”. *HP Labs* (basado en el análisis de los vínculos o links a páginas Webs externas que se incluyen en los blogs).

En este epígrafe hemos visto como en Internet las redes se configuran en torno a unos nodos principales o *hubs*, fenómeno que pudiera deberse a la unidireccionalidad de los links y al peso de la centralidad y popularidad de algunos individuos o grupos. Pero también que existen trabajos en los que se afirma que la centralidad en este canal no es relevante para la transmisión de tendencias o cómo los vínculos se multiplican enormemente en una red social en Internet como Facebook, o cómo la adaptación temprana de una idea en la Red junto con la interrelación entre comunidades, supone un indicador del volumen de difusión que alcanzará “físicamente”⁴⁶.

En los próximos epígrafes trataremos de acercarnos al comportamiento de las redes que se forman en Facebook y Twitter, tratando de establecer comparaciones entre sus conductas. El conocimiento de la forma en que estas redes se comportan es de utilidad para la validación de nuestra hipótesis, ya que podremos aprender el modo en el que en ellas suceden aspectos relevantes como el contagio o las cadenas de difusión, de forma que pueda quedar evidenciado que las características de su estructura red son suficientes para poder llevar a cabo una investigación a partir de los datos que extraemos de ellas.

2.4.2. Facebook y Twitter, nuevas redes

Una de las cuestiones que nos planteamos en este trabajo es comprobar si las redes sociales que se estructuran en Internet se comportan de la misma manera que las redes sociales en el mundo real, es decir, si nos conectamos y nos contagiamos bajo las mismas reglas.

Internet ha provocado cambios en las formas de las redes sociales, entre las que destaca la derivada de su capacidad ilimitada para generar nuevos vínculos, lo que Christakis y Fowler, designan como “enormidad”, refiriéndose a la vasta

⁴⁶ Esta revelación corrobora la capacidad de las Redes Sociales en Internet como herramientas de predicción de tendencias sociales, tal y como señala el título del estudio del Colbauhg y Glass.

magnitud de nuestras redes y del número de personas a las que se puede llegar. (Christakis, 2010, p.283)

Sabemos que en Facebook, por ejemplo, podemos llegar a tener miles de “amigos”. El término “amigo” en Internet se extiende en significado ya que con él agrupamos tanto a aquellos con los que nos unen lazos fuertes como débiles e incluso a contactos que ni siquiera conocemos y a los que añadimos a nuestra red por motivos profesionales, necesidad de información sobre temas de interés, o, simplemente, porque es “amigo de nuestros amigos”. Nuestra red se extiende, no aumentamos el número de personas con las que mantenemos una relación estrecha pero sí nuestro número de contactos a la vez que nos permite retomar conexiones perdidas con el paso de los años y que de otra manera no se celebrarían, entre otras causas, por dificultad geográfica.

Así, los “amigos” que se producen en las redes sociales se distinguen de nuestras amistades reales porque *tienden a ser acumulativas (en el mundo virtual la gente tiende a añadir conexiones y no cortarlas) y la naturaleza de la interacción se ve fuertemente influida por el medio (pequeñas llamaradas de actividad en lugar de conversaciones sostenidas, por ejemplo)*. (Christakis y Fowler, 2010, pp.284-285) El resultado es que al ampliar nuestra red hasta desconocidos que antes no existían se produce un incremento en la cantidad de información compartida.

En lo que se refiere al modelo de difusión en las redes sociales (a partir de aquí nos referiremos a los servicios de Internet como “redes sociales”), podemos avanzar que en Internet la difusión de información en cascada no se produce en largas y ramificadas cadenas sino que obedece a un patrón de difusión caracterizado por una larga escala de colisiones de cadenas cortas.

Las redes ofrecen nuevas vías para la influencia y el contagio social. Si como hemos anotado en epígrafes anteriores la frecuencia del contacto no es tan importante en la difusión de normas o conductas sociales, y, por tanto las ideas sobre comportamiento pueden difundirse incluso en ausencia de contacto personal, y,

además, hasta los amigos de los amigos de nuestros amigos –aunque estén a cientos de kilómetros- pueden influirnos, las redes sociales que se establecen en Internet como Facebook y Twitter se muestran como el entorno adecuado para que las ideas se propaguen fácilmente y contagien a un gran número de personas en un breve espacio de tiempo.

2.4.3. El contagio en Facebook

Facebook es una red social abierta a todos los individuos con el único requisito de disponer de una cuenta de correo. Las relaciones de “amigos” son recíprocas, puedes solicitar la amistad de quién desees pero éste tiene que aceptar tu solicitud para establecer la relación. En los perfiles personales se encuentra, entre otras, la opción *Wall* –muro- en la que los usuarios publican la información y que otros pueden comentar o compartir -difundir- entre sus amigos. El espacio común en el que se sigue la información emitida por nuestros “amigos” se llama *News Feeds* (de aspecto similar a la página Home de Twitter). Una de sus herramientas más válidas para la difusión de información es la opción *Share*, por medio de cual un contenido recibido de un usuario de nuestra red puede ser compartido con el resto de los usuarios de la misma, de forma que los usuarios de cada una de sus redes particulares puede acceder a él y de igual modo, volverlo a difundir.

Ilustración 4. Imagen de una red de usuarios de un particular en Facebook, realizada mediante la herramienta Touchgraph



En la ciencia de la sociología la red personal, o *personal network*, de un individuo es un objeto básico de estudio, centrándose para ello y en su mayoría, en su tamaño (número de contactos) y en su composición (estableciendo categorías como familiares y compañeros de trabajo). En Facebook, además de estas medidas de análisis, una red personal puede observarse desde una dimensión distinta: la manera en que un individuo divide su atención entre los contactos de su red, tomando en cuenta tanto con cuáles mantiene una actividad concreta y el tiempo que dedica a cada una de ellas⁴⁷ (Backstrom, Bakshy, Kleinberg, Lento y Rosenn, 2011).

⁴⁷ Se rastrearon las siguientes modalidades de actividad que los usuarios de Facebook pueden mantener con los individuos de su red: mensajes privados, comentarios a contenidos publicados, comentarios en el “muro”, visitas al perfil público y visionado de fotografías.

En su investigación, Backstrom et al. (2011) hallaron los siguientes comportamientos en las redes que se forman en Facebook⁴⁸:

- Existe una alta correlación entre el tamaño de la red y el nivel de actividad. Sin embargo, ambos parámetros no son suficientes para predecir los cambios de atención en el tiempo.
- Las actividades de comunicación se establecen preferentemente con los contactos principales de un usuario determinado, mientras que las de simple observación se dispersan entre el resto de contactos⁴⁹.
- Mientras que las actividades de observación se mantienen constantes a cualquier edad, a medida que esta aumenta disminuyen las actividades de comunicación.
- La actividad es mayor, en general, entre usuarios del mismo sexo.
- El alto nivel de actividad está asociado con mayores niveles de estabilidad de la misma en el tiempo.

El balance de atención (actividad en el tiempo) en Facebook, según el trabajo de Backstrom et al. (2011), es una propiedad relativamente estable en una red personal mientras que presenta variaciones entre grupos de personas diferentes (edad, género, etc.) y sus modalidades de atención.⁵⁰

Lazlo Barabási⁵¹ probó que la formación de redes en Internet no es aleatoria como en un principio se había pensado, sino que obedece a la “ley de potencias”. Profundizando en la formación de redes en Facebook, Chang, Rosenn, Backstrom y

⁴⁸ Comparan datos de actividad en usuarios de la red en dos periodos distintos en un mismo año (enero de 2010 y octubre de 2010). El rango de edad elegido es entre 13 y 60 años.

⁴⁹ Dichas actividades pueden clasificarse en dos grupos: comunicación, que contiene aquellas en las que el iniciador y el objetivo son conscientes de la acción (mensajes personales), y observación, que recoge las actividades en las que sólo el iniciador tiene consciencia de ello (visitas al perfil).

⁵⁰ Los cambios en la atención también se observan en Twitter: cuando el número de relaciones que se establecen en esta red supera el límite biológico definido por la hipótesis de Dunbar (160), la atención se limita y se priorizan internamente las interacciones. (Goncalves et al., 2011)

⁵¹ BARABASI, Lazlo y ALBERT, Réka (1999). "Emergence of scaling in random networks". *Science* 286: 509–512

Marlow (2010, p.1) hallaron, coincidiendo con trabajos anteriores, que la etnicidad en Facebook es selectiva en sus interacciones: *“We also find that different ethnic groups relate to one another in an assortative manner, and that these groups have different profiles across demographics, beliefs, and usage of site features.”*

Christakis y Fowler, siguieron durante un periodo cercano a los cuatro años a un grupo aproximado de 1.700 estudiantes conectados en Facebook para tratar de analizar como sucede el contagio en esta red, analizando la música, las películas y los libros más listados en los perfiles de los alumnos. Sus conclusiones fueron las siguientes (Christakis y Fowler, 2009):

- La influencia existe: *“When we want to figure out why someone likes this musician or that movie - out of all the musicians and movies in the world - a good place to start is by looking at the tastes of the people to whom he or she is connected.”*
- La diffusion de la influencia se sucede de igual manera que en el mundo real: *“The spread of influence online works in much the same way that it has worked offline for hundreds of thousands of years. In the past, present, and future, our closest connections are the ones that matter most.”*

En los detalles de su seguimiento se desprenden otras observaciones que son de interés para nuestro estudio ya que manifiestan algunas singularidades en el comportamiento del contagio:

- La red de películas está polarizada mientras que la de gustos musicales no lo está.
- Ciertas difusiones en la red no obedecen a ningún patrón de influencia conocido. Es el caso de los gustos de lectura, en el que los autores argumentan que la carencia de patrón se deriva del hecho de que la lectura es un hábito “solitario”, mientras que la música y las películas forman parte de lo que empleamos para divertirnos y, por tanto, compartimos más a menudo.

- No es sólo con quién estamos conectados lo que importa si no cómo estamos conectados. Para la difusión de ciertas ideas o gustos por la red, la centralidad no funciona y son emitidos en la periferia de la misma: *“Pulp Fiction was listed by more “popular” people at the center of the network, while Love Actually was preferred by the wallflowers on the periphery.”*

Ilustración 5. Difusión de gustos musicales en una red de amigos en Facebook, de Christakis y Fowler, (2009)

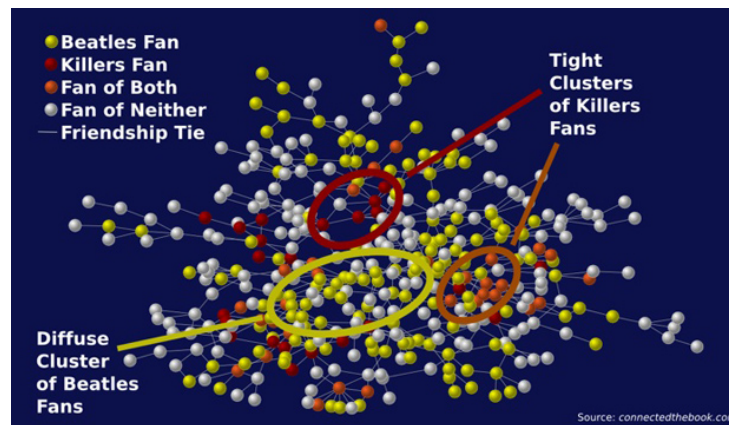


Ilustración 6. Difusión de gustos cinematográficos en una red de amigos en Facebook, de Christakis y Fowler, (2009)

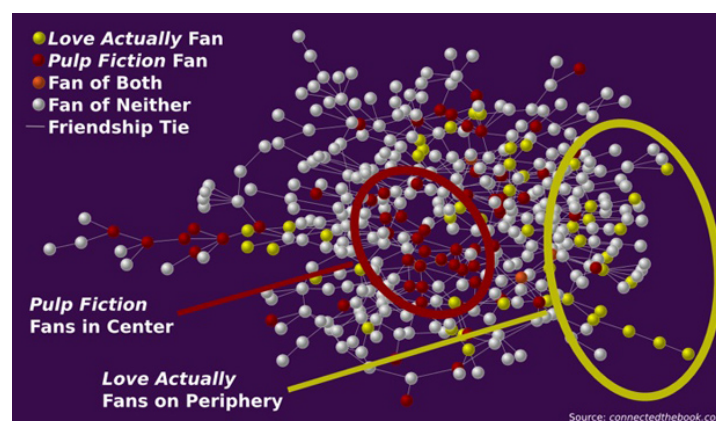
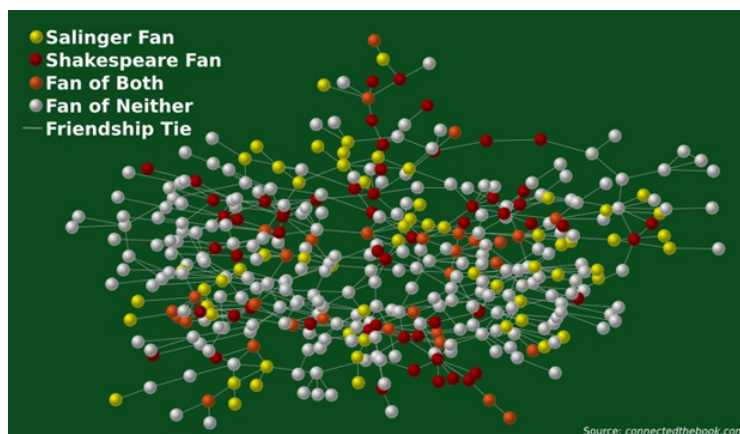


Ilustración 7. Difusión de gustos literarios en una red de amigos en Facebook, de Christakis y Fowler, (2009)



En el artículo “Friendships in the digital age” de *Oxford Analytica* (2010) se asegura que a pesar de Internet y de la creencia de que las redes sociales no tienen límite, la escala de nuestros círculos sociales permanece igual que siempre. Se apunta que históricamente las comunidades se formaban con alrededor de 150 personas. Dicho número coincide con el conocido como “Número de Dunbar”⁵² e indica no el límite pero sí el número de relaciones personales que un individuo puede mantener de forma recíproca. Se afirma en dicho artículo que nuestro mundo social consiste en una serie de capas concéntricas (5, 15, 50 y 150 contactos), que cada capa incorpora más individuos pero, a la vez, se deteriora la calidad de la relación entre ellos. Así, se argumenta que la confianza viene generada por el contacto o conocimiento directo, lo que supondría para el caso de la investigación que nos ocupa, asegurar que en las redes sociales en Internet, con cientos o miles de amigos, en las que el contacto es la mayoría de las veces única y exclusivamente online e incluso indirecto (sin

⁵² Según el antropólogo Robin Dunbar, 150 es la cantidad de individuos que pueden desarrollarse plenamente en un sistema determinado. Está relacionado con el tamaño del neocórtex cerebral y su capacidad de proceso.

La replicación de la hipótesis de Dunbar en la red Twitter será expuesta más adelante, en el epígrafe dedicado al contagio en esta red.

conversación bidireccional), no podríamos confiar en las ideas o mensajes lanzados por estos.

La opinión vertida en este artículo no ha tenido en cuenta que en Internet, por sus propias características de universalidad, heterogeneidad y deslocalización, es posible establecer vínculos con más de una comunidad⁵³, haciendo de puente (lazo débil) entre aquellas a las que se pertenece. Los individuos puente hacen posible que la información se transmita (y contagie) entre ellas, realidad que de otra forma no sucedería. Venimos a decir, que si bien entendemos la veracidad lanzada por el “Número Dunbar” en lo que a relaciones bidireccionales se refiere, las cuales podríamos asignar a lo que en teoría de redes se denomina como “vínculos fuertes”, siguiendo de forma única la teoría de Dunbar, estaríamos dejando a un lado otras teorías de gran peso en el análisis de redes sociales como la ya mencionada de la importancia de los vínculos débiles expuesta por Granoveter (1972), la cuál consideramos de vital importancia en este estudio, dada la naturaleza, espacios y condiciones en los que las Redes Sociales en Internet se suceden. Los vínculos fuertes tienden, por naturaleza, a la homofilia, mientras que los débiles sirven de puente entre polarizaciones o tendencias diferentes.

Los modelos de difusión han sido utilizados para explicar el fenómeno del contagio, la forma en que este ocurre en una red social determinada. Tradicionalmente, el método empleado en dichos modelos ha sido el heredado de los modelos epidemiológicos en los que, en su mayoría, se parte de un evento aislado y se exploran las condiciones bajo las cuales este evento desencadena una cascada global. Asumen que unos pocos nodos ponen en marcha largas cadenas de difusión que terminan en una extensa cascada global⁵⁴. Sin embargo, Sun, Rosenn, Marlow y Lento (2009) pusieron en evidencia que el modelo de difusión en Facebook difiere de

⁵³ La diferencia entre redes sociales y cualquier otro tipo de red no social es que está dividida en comunidades. (Newman, 2003: 8).

⁵⁴ ANDERSON, R., y MAY, R. (1991). “Infectious Diseases of Humans: Dynamics and Control”. New York, NY: Oxford University Press.

los modelos clásicos en lo que se refiere al peso de los influyentes y a la longitud de las cadenas “infectadas”. El estudio “Gesundheit! Modeling Contagion through Facebook News Feed”, en el que se vigila la difusión de las “Fun Pages”, páginas concebidas como perfiles distintos al resto, diseñados a medida para empresas, marcas, celebridades, políticos, etc. con el fin de representarse a sí mismos realizado en el transcurso del año 2008, asegura que mientras que tradicionalmente se admitía que un evento comienza con un reducido número de nodos iniciadores que son capaces de afectar a los vecinos más vulnerables, en Facebook los contagios que se convierten en cascadas globales se dan en eventos que comienzan en un gran número de nodos que ponen en marcha un gran número de cadenas cortas, cada una de las cuales colisiona con rapidez en una sola estructura de gran tamaño⁵⁵. Los investigadores también hallaron que en la longitud de la cadena no influyen las características demográficas de sus iniciadores siendo el aumento del número de páginas personales en las que el evento aparece en los *mensajes* que publican en sus *wall's* o muros y que se comparten automáticamente en los *News Feed* de la home de la red, la única alternativa válida para incrementar la longitud de la cadena⁵⁶. Por tanto, concluyen, no se puede identificar a los iniciadores como lo más importante en un proceso de difusión cultural, ya que en redes sociales como Facebook una gran cantidad de individuos tienen la posibilidad de desencadenar una cascada global de gran tamaño: *“Although Facebook diffusion chains are often extremely long (chains of up to 82 levels have been observed), they are not usually the result of a single chain-reaction event. Rather, these diffusion chains are typically started by a*

⁵⁵ El estudio se llevó a cabo durante seis meses, con los datos recogidos del 262.985 páginas de Facebook, fijándose en la propagación de los *news feed* y analizando detalladamente, 179.010 nodos iniciadores de las cadenas de difusión.

El fenómeno de las cascadas sociales que se dan en el contagio, aunque referido al comportamiento de difusión de de vínculos a fotografías publicadas en la red Flickr, fue estudiado por CHA, Meeyoung et al. (2008): *“We found that social cascades are occurring in the spread of photos though the Flickr social network. (...) We found the basic reproduction number of popular photos to be between 1 and 190. This is much higher than very infectious diseases like measles, indicating that social networks are efficient transmission mediums and online content can be very infectious.”*

⁵⁶ La difusión de una Fun Page ocurre cuando: 1. Un usuario se hace seguidor de una página, 2. Esta acción se emite en los News Feeds de sus amigos, y 3. Uno o más de esos amigos, lo ve y decide convertirse también en fan de dicha página.

substantial number of users. Large clusters emerge when hundreds or even thousands of short diffusion chains merge together (...) there is no meaningful evidence that a start node's maximum diffusion chain length can be predicted with the user's demographics or Facebook usage characteristics (including the user's number of Facebook friends)." (Sun et al., 2009, p.1)

Debido al tipo de conectividad que Facebook provee y la facilidad para convertirse en fan de una página, la longitud de la cadena de difusión puesta en marcha por un nodo iniciador pueden llegar a ser extremadamente largas en comparación con estudios referidos al mundo real. Además, las características demográficas de los nodos iniciadores, que ascienden a un 18% en la mayoría de los grupos analizados a diferencia de las condiciones de iniciación que se han visto en la literatura tradicional, no tienen un impacto significativo en la longitud de la cadena, siendo el incremento de la probabilidad de que la acción de convertirse en fan de una página aparezca en los News Feeds de otros usuarios, la única manera de incrementar la longitud de la cadena de difusión (Sun et. al 2009, pp.7-8). Por tanto, argumentan, no es fácil concluir que los iniciadores son lo que más importa o afecta en un ciclo de contagio: *"After looking at the distribution of "start nodes" and "follower nodes" in these clusters, we find no evidence to support the theory that just a few users are responsible for the popularity of Pages."* (Sun et. al 2009, p. 5)

No obstante, los autores advierten que los resultados de su investigación no podrían directamente extrapolarse a otros espacios, debido a la especificidad de la conducta social que ocurre en esta red, propiciada por las diferencias en las características de su "población", las especiales normas sociales que dictan la interacción y la influencia en ella, y la forma en que un individuo se relaciona con las Fan Pages en contraste con otras formas de contenido.

Un dato revelador, por su relación con las premisas definidas en el análisis de redes sociales para los "early-adopters", los iniciadores de las cadenas de difusión en Facebook, al menos en la *Fan Pages* objeto del estudio de Sun et al. (2009), es el hecho que indica que estos iniciadores se convirtieron en usuarios de la red antes que

los “no iniciadores”, siendo esta característica una constatación de que adoptaron prontamente esta Red Social y, por tanto, pertenecer a la categoría social de adaptadores tempranos. Así, este hallazgo coincide con las teorías que explican que los “early-adopters” poseen un grado destacable de influencia sobre el resto. Además, estos iniciadores son usuarios más activos en la Red, produciendo más cantidad de contenido y con una predisposición a buscar nuevas Fan Pages.

Un estudio que cubría 28 países y 17.000 usuarios de Internet, fechado en 2008, sobre la influencia en los *social media* de la consultora McCann revela, aunque asuma que ahora todos podemos influir en los demás, la aparición de una nueva clase de influyentes a los que denomina “super-influyentes”. Individuos que gracias a los social media y a las herramientas que proveen para crear y compartir juicios, superan el grado medio de la influencia difundiendo su opinión más lejos y más ampliamente. Son personas inteligentes, con un alto nivel de educación, profundo conocimiento de Internet y ordenadores, generan y comparten contenido, y se ajustan al perfil de *early-adopters* (adaptadores tempranos) dispuestos a probar nuevos productos, asumir riesgos y compartir sus opiniones con amigos. La consultora afirma que la actividad de esta clase de “super-influyentes” fue determinante en la victoria electoral de Barack Obama en las elecciones de 2008 en Estados Unidos: *“It has been widely credited that the creators or creative class, as they have been tagged, helped Obama get the democratic party nomination. The creators through prolific blogging, twittering, social networking and content creation generated massive influence helping to raise funds, getting people out to vote and changing the opinions of mainstream media.”* (Universal McCann, 2008, pp.55-56)

Sin duda, en el estudio del fenómeno de la viralidad, esto es qué características debe poseer un objeto para extenderse y adoptarse (contagiarse) junto con las características de la red en la que es capaz de propagarse, un análisis de la mayor relevancia lo constituye el conocimiento y modelado del número y tamaño de las comunidades que infecta y de las vías por las que es capaz de viajar entre ellas. Granoveter (1973) ya evidenció la importancia de los lazos débiles como nexos o

puentes entre grupos diferentes, artífices del salto del contagio. En el estudio de Sun et al. (2009), citado anteriormente, se descubrió que la confluencia de eventos aislados que concurren en una gran cascada es consecuencia de la unión de dos o más cadenas de difusión debido a que un usuario se convierte en fan de una página después de haber visto esta actividad en dos amigos pertenecientes a diferentes cadenas de difusión. Dichas actividades, que aparecen en los *News Feeds* tomarían la forma de “recomendaciones” implícitas que viajan entre diferentes comunidades en la red, dibujando un mapa que contiene un número de usuarios de diferentes grupos, cuyo volumen sirve para determinar si aquello que se propaga ha adquirido la condición de viral.

Coincidiendo con esta afirmación, el hecho de que un usuario se convierte en fan de una página después de haber visto esta actividad en dos amigos pertenecientes a diferentes cadenas de difusión, se encuentran las conclusiones de Ugander, Backstrom, Marlow y Kleinberg (2012) que atribuyen a la diversidad de la red un rol crucial en la propagación. Los autores advierten que en vez de considerar el número de contactos de un individuo (tamaño de la red) como un parámetro crucial en el análisis del contagio social, debería considerarse el número de contextos sociales en los que éste se desenvuelve:

“More broadly, across a range of further domains, these findings suggest an alternate perspective for recruitment to political causes, the promotion of health practices, and marketing; to convince individuals to change their behavior, it may be less important that they receive many endorsements than that they receive the message from multiple directions.” (Ugander et al., 2012, p.4)

Mientras que los modelos tradicionales de análisis del contagio social se han basado históricamente en las analogías físicas con el contagio en epidemias biológicas, en las que las probabilidades de que un individuo se infectara o contagiara crecían monótonamente con el tamaño de sus contactos (o número de

individuos infectados con los que ha mantenido contacto), vigilando el crecimiento de Facebook, dada su particularidad como insólito ejemplo de proceso social de adopción global, encontraron que la probabilidad de contagio está estrechamente controlada por el número de componente conectados con un individuo en la red y no tanto por el tamaño de la misma:

Surprisingly, once this “structural diversity” is controlled for, the size of the contact neighborhood is in fact generally a negative predictor of contagion.” (Ugander et al., 2012, p.4)

En 2007, Facebook tomó la decisión trascendental de abrir su plataforma para el desarrollo de aplicaciones independientes que funcionaran sobre ella, sus afiliados y sus actividades. Después de eso, muchas han sido las aplicaciones que se han diseñado y desplegado en esta plataforma. Tal notoriedad acompañada de la del análisis ampliado que Facebook permite acerca de los usuarios que “ceden”⁵⁷ los datos de su perfil personal al adoptar una aplicación desarrollada para funcionar en Facebook e incluida en su directorio de aplicaciones, han amplificado el número y el tipo de análisis que pueden obtenerse en esta red. Centrándonos en el análisis político que nos interesa, cabe decir que el desarrollo de las páginas Fun Page, destinadas a celebrities de todo tipo y condición, supuso el uso masivo de éstas por parte de candidatos en campaña electoral. Dichas páginas, debido a su carácter “público”, son susceptibles de seguimiento indiscriminado (útil para la investigación) de lo que acontece en ellas. Recientemente, no sólo investigadores sino aplicaciones comerciales dedicadas a la captura de su total actividad han saltado a la escena, valga citar la herramienta de análisis Smetrica.com (<http://www.smetrica.com>), desarrollada por dos estudiantes de la isla de Gran Canaria (España), la cual permite

⁵⁷ La información referida a un usuario es restringida al conocimiento público o universal (filosofía contraria a la de la red Twitter). La privacidad de los perfiles publicados en Facebook supone que, dado un perfil de usuario, son accesibles simples datos acerca de la edad, sexo y número de amigos, y no más (un usuario o un aplicación de minería de datos no puede acceder a lo que publicas y, por tanto, entre quién y cómo lo publicas). La bien conocida privacidad de Facebook puede saltarse cuando un usuario cede sus datos al aceptar, inconscientemente, que las aplicaciones que adopta cotilleen profundamente en él mismo y en su red.

la visualización de la actividad que ocurre en un determinado perfil en una fecha indicada⁵⁸.

La difusión en el uso de la aplicación⁵⁹ de notas *Notecentric*⁶⁰ entre usuarios de Facebook, fue estudiada por su propio desarrollador, Hung Truong. La visualización de la red que hacía uso de la misma descubrió que, junto con los usuarios únicos o aislados que la utilizaban (59%) y no lo comentaban con otros, convivían dos grandes grupos conectados: usuarios de diferentes grupos que propagaban la aplicación a otros (10%), y un grupo homogéneo de amigos que la compartían – posiblemente compañeros de clase- (25%). El autor concluyó que la aplicación no era suficientemente viral, al conseguir, probablemente, que la mayoría de sus adeptos (usuarios únicos) llegaran a ella a través del directorio de aplicaciones de Facebook y no a través del *word of mouth* (el boca a boca). No obstante, esta fragmentación dibujada en su red de usuarios, pone de manifiesto que en los fenómenos de difusión la red está formada por un conjunto de grupos o comunidades o sub-redes. Ser capaz de saltar de una a otra supone la diferencia entre alcanzar o no la viralidad. (Truong, 2009)

A tenor de lo expuesto en los análisis reflejados arriba, vemos que la influencia sí existe y es mayor entre nodos cercanos pero los influyentes no son determinantes para una difusión global en cascada. Entonces, nos hacemos las siguientes preguntas:

- Siguiendo la teoría de Barabási, en la que se asegura que los *hub*'s son imprescindibles para el sostenimiento de la comunidad ¿Existen

⁵⁸ La autora del presente trabajo ha realizado en octubre de 2010 un estudio de predicción electoral sobre las Elecciones Generales 2011 en España, mediante el análisis de los datos obtenidos con esta herramienta (smetrica.com). Los resultados obtenidos están desviados de la realidad (resultado electoral). La autora esgrime que tal volumen de desviación pudiera deberse a las características del perfil del usuario de Facebook, un tanto alejado del perfil más “serio” de los usuarios de Twitter.

⁵⁹ De los 648 de Notocentric, 440 la habían adoptado después de que uno de sus amigos en Facebook lo hubiera hecho con anterioridad.

⁶⁰ Aplicación de notas compartidas en Facebook, desarrollada usando la plataforma *Facebook Developer*.

nodos en Facebook que alcanzan gran popularidad por su gran actividad en temas concretos y por su creciente número de seguidores que los encumbran, en cierto modo, como “gurús” que suministran información fiable y en los que, debido a su calidad y visibilidad en la red, se deposita confianza y sus opiniones e ideas se toman como verdades que creer y por tanto, re-difundir su mensaje pudiendo influir en otros?

- Granovetter propuso que la confianza en un líder depende de si existen contactos directos o indirectos que pueden asegurarnos dicha confianza (Granovetter, 1973, p.13), ¿la popularidad de algunos individuos en Facebook les convierte en una especie de líderes de forma que, por un lado, se confía en ellos y, de otro, pueden hacer de puente o intermediario entre un líder superior y el resto de la red?

Consideramos que en el caso que nos ocupa en este trabajo, sería del mayor interés y necesidad contar con estudios que tuvieran en cuenta el papel que juegan los individuos que dentro de una red son respetados y seguidos por su experiencia en un tema concreto (política, en nuestro caso), debiendo rastrear el alcance de la difusión de sus mensajes (número de infectados y tiempo) en comparación con otros individuos, así como la naturaleza semántica de las opiniones o reacciones que provoca en los alcanzados. Juntamos aquí varias aspiraciones difíciles de alcanzar en el análisis de una cibercampaña: la determinación de los perfiles más influyentes (no sólo basado en número de seguidores sino en qué proporción sus mensajes son seguidos o retuiteados, la identificación y seguimiento de las cadenas de difusión de mensajes (cuyos actores pueden o no coincidir con los perfiles influyentes, identificando puentes débiles y actores desconocidos capaces de super-difusión), y qué o cuáles términos obtuvieron una mayor propagación en la red (en este ítem entraría en escena el *sentimental analysis*, cuya dificultad y/o carencia será discutida en páginas siguientes).

2.4.4. El contagio en Twitter

Twitter es un servicio de *microblogging*⁶¹ en el que, a diferencia de la mayoría de redes sociales en Internet, las relaciones que se establecen entre sus usuarios no están obligadas a ser recíprocas, es decir, cualquier individuo puede iniciar el seguimiento de otro (*following*) sin que este segundo tenga que autorizarle - omitiendo perfiles privados o cerrados- o añadirle como amigo, y a su vez, si alguien sigue a un individuo, éste no está forzado a devolverle el seguimiento. Esta, junto a la tipología de los mensajes (*tweet*) que circulan por este medio en su mayoría de carácter noticioso, han propiciado que se genere un debate sobre si Twitter es una red social o un medio de comunicación social. Twitter es una red social en cuanto conecta a personas y a la vez, un medio de comunicación social, debida esta dualidad quizás a sus características de difusión de emociones, noticias y otro tipo de expresiones (limitación del mensaje expresado a 140 caracteres, no exigencia de seguimiento recíproco entre usuarios, y sus herramientas propias de etiquetado de la información que circula, entre otras).

Más allá de su clasificación, lo que Twitter si ha demostrado es una capacidad asombrosa para difundir información entre sus millones de usuarios. Las herramientas de que surte para facilitarla son los *retweet*⁶² (RT), los *hashtag*⁶³ (#) y derivados de estos, los *Trending Topics*⁶⁴ o temas de máxima actualidad y actividad en la red.

⁶¹ Servicio que permite a sus usuarios enviar y publicar mensajes breves (alrededor de 140 caracteres), generalmente de sólo texto. Las opciones para el envío de los mensajes varían desde sitios web, a través de SMS, mensajería instantánea o aplicaciones *ad hoc*.

⁶² ReTweet (RT). Mensaje publicado por un individuo que es replicado por otro para compartirlo con sus seguidores. Los RT son útiles para propagar la información en esta red.

⁶³ Forma de etiquetar una palabra de forma que pueda luego ser buscado por otros. Su versatilidad les ha convertido en generadores de conversaciones multitudinarias sobre temas concretos

⁶⁴ Los *trending topics* tienen su origen en los *hashtags*, etiquetas precedidas por una almohadilla que sirven para organizar tuiteos sobre un tema concreto en Twitter. Lo que en principio era un ránking de los nueve *hashtags* más utilizados terminó por convertirse en una radiografía de los temas más hablados, aunque no fueran *hashtags*. http://es.wikipedia.org/wiki/Trending_topic. (Consultado el 05/04/2011)

En dicha capacidad de difusión, la posibilidad de mantener conectados a cientos de usuarios juega un papel decisivo y, a la vez, diferenciador de otro tipo de redes sociales. Aunque el objeto de su estudio fue determinar si la hipótesis conocida como el “Número Dunbar” se replicaba en una red social como Twitter, Gonçalves, Perra y Vespignani (2011) generaron una red de conversaciones (cada conversación en twitter adquiere la forma de un árbol de *tweets*) que conectaba alrededor de 1,7 millones de individuos⁶⁵.

Ilustración 8. Red de conversaciones en Twitter de Gonçalves et al. (2011)

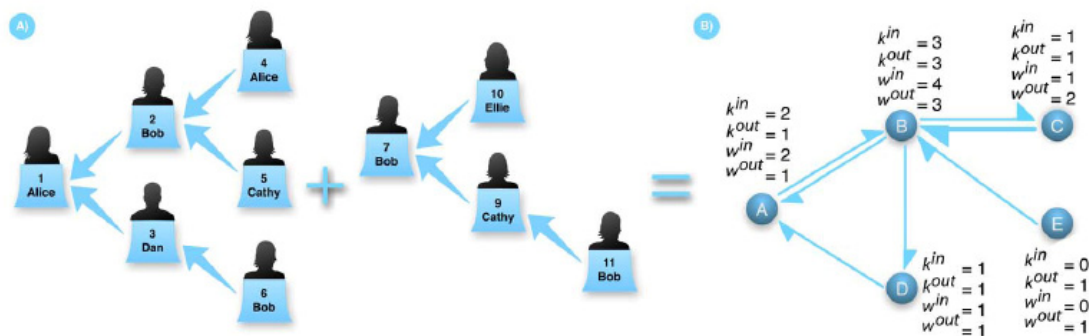


Figure 1. Reply trees and user network. A) The set of all trees is a forest. Each time a user replies, the corresponding tweet is connected to another one, resulting in a tree structure. B) Combining all the trees in the forest and projecting them onto the users results in a directed and weighted network that can be used as a proxy for relationships between users. The number of outgoing (incoming) connections of a given user is called the out (in) degree and is represented by k^{out} (k^{in}). The number of messages flowing along each edge is called the degree, w . The probability density function $P(k^{out})$ ($P(k^{in})$) indicates the probability that any given node has k^{out} (k^{in}) out (in) degree and it is called the out (in) degree distribution and is a measure of node diversity on the network.
doi:10.1371/journal.pone.0022656.g001

En el estudio “What is Twitter, a social network or a news media?” (Kwak, Lee, Park y Moon, 2010) sus autores rastrearon la esfera Twitter completa en julio de 2009, obteniendo 41,7 millones de perfiles de usuarios, 1,47 billones de relaciones sociales, 4.262 *trending topics* y 106 millones de *tweets*. Las principales conclusiones a las que se llegaron fueron que se cumple escasamente la “ley de potencias” en la distribución de seguidores; que existe un grado bajo de reciprocidad en las relaciones; que la mayoría de los temas se corresponden con titulares de noticia o similares; que el concepto de popularidad se obtiene tanto por número de

⁶⁵ Analizaron cerca de 380 millones de *tweets* de los que extrajeron 25 millones de conversaciones.

seguidores como por la popularidad de un *tweet*; y que tras el primer *retweet*, un mensaje es re-difundido por otros a sus seguidores instantáneamente.

Junto con estas conclusiones, su análisis desveló información detallada sobre cómo funciona Twitter, cómo es el comportamiento de sus usuarios y cómo se procede a la difusión de la información. Listamos a continuación los datos más relevantes:

- La formación de la cadena
 - Sólo hay 40 usuarios con más de un millón de seguidores, siendo todos ellos celebridades (artistas, periodistas y políticos). Popularidad.
 - Sólo el 22% mantiene una relación recíproca. Baja reciprocidad.
 - El 67% de usuarios no es seguido por ninguno de los usuarios a los que él sigue. Twitter como fuente de información.
 - 4,12 es el grado de separación entre un usuario y su usuario objetivo (menor que los 6 grados de Milgram)
 - En las relaciones recíprocas se comprueba la tendencia a contactar con gente similar. Homofilia.
- Temas que marcan tendencia
 - Más de 8 millones de usuarios participan en los *trending topics* y cerca del 15% lo hace en más de 10 en un mes.
 - El periodo de tiempo en el que un *trending topic* está activo no supera la semana.
 - Los titulares de noticias son *trending topics* de corta duración y los relacionados con el deporte, las ciudades y las marcas adquieren una duración mayor. Un tercer grupo lo forman los *trending topic* efímeros.
- El impacto del *retweet*

- Un usuario puede determinar qué información es importante y hacer un *retweet* de la misma, posteriormente el colectivo decidirá si la información contenida en el *tweet* original lo era o no. Inteligencia colectiva.
- La mitad de los *retweets* se producen en el intervalo de una hora. El 75% durante un día y sólo el 10% durante un mes. Instantaneidad de la noticia.
- La gente sólo hace *retweet* de un pequeño grupo de personas. Favoritismo.
- Los *retweets* alcanzan una gran audiencia y se difunden rápidamente. Poder de difusión.

Más allá de las intenciones originales del estudio, catalogar a Twitter como red social o como fuente de información, dos son los aspectos que nos resulta interesante destacar por su implicación con el objeto de nuestra investigación:

- Su capacidad de contagio, gracias a la visibilidad del *retweet*.
- La escasa influencia de la centralidad, medida en popularidad o número de seguidores, en la difusión.
- El papel relevante de los “lazos débiles”, individuos que actúan como puente no ya entre nodos, sino entre diferentes comunidades o redes al alcance.

Sobre el impacto de los *retweets* ha investigado Dan Zarella. En su informe “The science of retweets”, concluyó que aunque los usuarios con más seguidores tienen más posibilidad de obtener *retweets* (por el número en sí, obviamente), algunos individuos son capaces de conseguir muchos *retweets* con un bajo número de seguidores, argumentando que este hecho se debe a que su contenido se considera interesante y contagioso. Otras conclusiones que se recogen en su estudio son las siguientes (Zarella, 2009, pp.5-22):

- Los usuarios *suggested* -recomendados para su seguimiento- por Twitter, aunque son las personas con mayor número de seguidores en esta red, no son los que consiguen más *retweets*.
- La presencia de un link al contenido en el mensaje del *tweet* puede incrementar su posibilidad de ser compartido por otros.
- Los link más cortos, realizados con “bit.ly”, “ow.ly”, o “is.gd”⁶⁶ son más compartidos que los hechos con “tinyurl”, quizás debido a su mínima longitud de caracteres.
- Las diez palabras que más se comparten son: *you, twitter, please, retweet, post, blog, social, free, media y help*. *You* indica que si “tú hablas conmigo”, más haré un retweet. *Retwet* indica su poder cuando “llamas a la acción”⁶⁷.
- Las diez que menos: *game, going, haha, lol, but, watching, work, home, night y bed*. Las actividades mundanas no son objetivo de *retweets*.
- Los *retweets* presentan mayor complejidad en sus sílabas y palabras, más caracteres de puntuación, y requieren mayor nivel cultural de los usuarios para ser comprendidos.
- La novedad del mensaje influye en el número de *retweets*.
- El contenido emocional influye menos que lo conceptual. El trabajo, la religión, el dinero y los medios de comunicación/ celebridades reciben más *retweets* que las emociones negativas, las sensaciones, las palabrotas y la auto-referencia.
- Los *retweets* se producen en mayor cantidad entre las 15:00 y la media noche.

⁶⁶ Debido a que los mensajes que se emiten en Twitter no pueden exceder de 140 caracteres, surgieron herramientas en Internet que convierten cualquier dirección Web en una “dirección corta”.

⁶⁷ Nótese que el estudio se refiere a términos en inglés. La autora desconoce investigaciones al respecto en lengua castellana.

El desarrollo de Internet ha cambiado el modo en que los “memes” se contagian: se difunden rápidamente, se exponen a más individuos y se superan las barreras impuestas por el límite de la memoria humana. Investigar el modo en que los *retweet* se producen se convierte en una herramienta de gran utilidad en la comprensión de los mecanismos de propagación de ideas en la red: *Not only can this information help us create more contagious Tweets, but many of the lessons learned through ReTweets will be applicable to viral ideas in other mediums.* (Zarrella, 2009, p.4)

Sobre el segundo aspecto que nos hemos marcado como relevante en el estudio del contagio en Twitter, referido al mayor o menor peso de la centralidad (medida como popularidad o número de seguidores), sin duda las conclusiones del estudio llevado a cabo por Cha, Mislove, Adams y Gummadi (2010) suponen un hallazgo relevante. Según las conclusiones a las que llegaron sus autores, los usuarios más populares que alcanzan altos números de seguidores (como es el caso de las *celebrities* como Britney Spears) no son necesariamente influyentes en su red, en términos de engendrar retweets o menciones. Midiendo número de seguidores, retweets y menciones como variables indicadoras de influencia en Twitter, *the top influentials* son generalmente, figuras públicas (ej. Barack Obama) y sitios Web (ej. CNN).

Entre los 6 millones de usuarios activos en Twitter, encontraron 233 usuarios a los que denominaron *all-time influentials* y monitorizaron sus puntuaciones de influencia durante los meses de enero y agosto de 2009. En aras de capturar la tendencia con más detalle, clasificaron al grupo de influyentes en tres grupos basados en el número de seguidores (indegree): los 100 más influyentes se corresponden con las principales agencias de noticias, los siguientes 90 eran principalmente *celebrities*, y el resto, una mezcla de figuras públicas y líderes de opinión que compiten con los medios (llamados *evangelistas*). La evolución de su influencia en el tiempo sucedió de la siguiente manera:

- Las agencias de noticias son los usuarios más retwiteados. Aparecen cuando generan contenido que se retuitea pero no les es fácil atraer audiencia por sí solas.
- Las *celebrities*, los más mencionados. Los usuarios normales sienten una enorme pasión por ellas.
- Los *evangelistas* incrementan su influencia con éxito. Dedicar esfuerzos a conversar con otros (responder a su audiencia) puesto que necesitan autopromocionarse.

También encontraron explicación para el hecho de que algunos usuarios normales u ordinarios consiguieran incrementar su influencia durante un corto periodo de tiempo. Se trata de usuarios que limitan sus tweets a un único tema y que propagan muy activamente información sobre protestas y noticias controvertidas. (Cha et al., 2010, pp.6-7)

“We found that indegree represents a user’s popularity, but is not related to other important notions of influence such as engaging audience, i.e., retweets and mentions. Retweets are driven by the content value of a tweet, while mentions are driven by the name value of the user. Such subtle differences lead to dissimilar groups of the top Twitter users; users who have high indegree do not necessarily spawn many retweets or mentions. This finding suggests that indegree alone reveals very little about the influence of a user.” (Cha et al., 2010, p.8)

Esta afirmación de que la popularidad no es sinónimo de influencia fue también constatada por Romero et al. (2010) en su estudio sobre cómo se propaga la información en Twitter: *“An important conclusion from the results is that the correlation between popularity and influence is quite weak, with the most influential users not necessarily being the ones with the highest popularity.”* (Romero, Galuba, Asur y Huberman, 2010-2)

¿Qué contenido recibe atención? Existe un consenso generalizado en considerar que la popularidad y la influencia de los usuarios son dos de los aspectos más relevantes en el proceso de transmisión de la información. Los autores miden la popularidad en número de seguidores que automáticamente reciben el contenido que un usuario genera, y la influencia depende no sólo del tamaño de su audiencia sino también del índice de pasividad - la pasividad de un usuario se mide por la dificultad que supone para otros usuarios influir en él.- de ésta a la hora de propagar el contenido que recibe.

Se capturaron, utilizando Twitter Search API⁶⁸, todos los tweets que contenían la cadena “http”⁶⁹ (evidencia un link o enlace a una página o dirección Web) durante un periodo de 300 horas continuadas, empezando el 10 de septiembre de 2009, acumulando un número final de 22 millones de tweets. Aplicando su propio algoritmo de influencia y pasividad, obtuvieron los siguientes descubrimientos:

- La lista con los 10 usuarios (usuarios que publicaron al menos 10 direcciones URL en el periodo de captura) más influyentes está dominada por agencias de noticias sobre política, tecnología y medios sociales (Internet)
- La lista con los 10 usuarios más pasivos (usuarios que siguen a muchos otros, pero que retuitean un porcentaje muy pequeño de la información que consumen) contiene *robots*⁷⁰ (agregan

⁶⁸ El API Search de Twitter es una API dedicada a efectuar búsquedas sobre los tweets recientes.

API o Interfaz de programación de aplicaciones (del inglés *Application Programming Interface*) es el conjunto de funciones y procedimientos (o métodos, en la programación orientada a objetos) que ofrece cierta biblioteca para ser utilizado por otro software como una capa de abstracción. Son usadas generalmente en las bibliotecas (también denominadas vulgarmente “librerías”). http://es.wikipedia.org/wiki/Interfaz_de_programaci%C3%B3n_de_aplicaciones. (Consultado el 12/07/2012)

⁶⁹ Hypertext Transfer Protocol o HTTP (en español protocolo de transferencia de hipertexto) es el protocolo usado en cada transacción de la World Wide Web. <http://es.wikipedia.org/wiki/Http>

⁷⁰ Robots de Internet, también conocidos como Robots Web, Robots WWW o *bots* simplemente, son aplicaciones de software que ejecutan tareas automatizadas a través de Internet. Típicamente, los robots realizan tareas que son a la vez simples y repetitivas estructuralmente, a una tasa mucho más alta de lo que sería posible para una única persona. El mayor uso de los robots es en *spidering web*, en

automáticamente keywords -palabras clave- o contenido de cualquier usuario de la red), cuentas o perfiles suspendidos (posiblemente de *spammers*), y usuarios que publican información muy a menudo.

- Los usuarios con mayor número de seguidores pero con escasa influencia. Son usuarios muy populares que tienen la atención de muchos otros pero no consiguen que sus mensajes se propaguen por la red debido a que su audiencia no los considera suficientemente importantes para reenviar a otros.
- Los usuarios con pocos seguidores pero con un alto grado de influencia. Entre ellos se encuentran usuarios que publican dibujos generados con la herramienta *twitdraw.com*, políticos locales y dibujantes o caricaturistas de políticos.
- La lista con las agencias o fuentes de noticias más influyentes, en la que se constata nuevamente que no se corresponden con aquellas que poseen mayor número de seguidores.

De sus conclusiones se desprende, además de que la popularidad no es sinónimo de influencia, que para que un contenido se propague por la red es necesario que otros lo difundan a su vez (audiencia activa) y que esto suele ocurrir cuando el contenido que se publica por primera vez es contenido considerado de valor o calidad: *“This is a reflection of the fact that for information to propagate in a network, individuals need to forward it to the other members, thus having to actively engage rather than passively read it and rarely act on it.”* (Romero et al., 2010, p.9)

el que un *script* automatizado obtiene, analiza y presenta información de servidores Web a varias veces la velocidad de un humano. Cada servidor puede tener un archivo llamado *robots.txt*, que contiene normas para ejecutar un *spidering* en ese servidor que el *bot* se supone que debe obedecer. http://en.wikipedia.org/wiki/Internet_bot. (Consultado el 27/04/2012)

Tabla 1. Tablas de relación usuarios-influencia extraídas del trabajo de Romero et al. (2010)

mashable	Social Media Blogger
jokoanwar	Film Director
google	Google
aplusk	Actor Ashton Kutcher
syfy	Science Fiction Channel
smashingmag	Online Developer Magazine
micellemalkin	Conservative Commentator
theonion	News Satire Organization
rwk	Tech/Social Media Blogger
breakingnews	News Aggregator

Table 1: Users with the most IP-influence (with at least 10 URLs posted in the period)

redscarebot	Keyword Aggregator
drunk_bot	Suspended
tea_robot	Keyword Aggregator
condos	Listing Aggregator
wootboot	Suspended
raybeckerman	Attorney
hashphotography	Keyword Aggregator
charlieandsandy	Suspended
ms_defy	Suspended
rpattinsonbot	Keyword Aggregator

Table 2: Users with the most IP-passivity

User name	Category	Rank by # followers	Rank by IP-influence
thatkevinsmith	Screen Writer	33	1000
nprpolitics	Political News	41	525
eonline	TV Channel	42	1008
marthastewart	Television Host	43	1169
nba	Sports	64	1041
davidgregory	Journalist	106	3630
nfl	Sports	110	2244
cbsnews	News Channel	114	2278
jdickerson	Journalist	147	4408
newsweek	News Magazine	148	756

Table 3: Users with many followers and low relative influence

User name	Category	Rank by # followers	Rank by IP-influence
cashcycle	Retweet Contest	153286	13
mobiliens	Retweet Contest	293455	70
jadermattos	Twitdraw	227934	134
jaum	Twitdraw	404385	143
robmillerusmc	Congressional Candidate	147803	145
sitekulite	Twitdraw	423917	149
jesse_sublett	Musician	385265	151
cyberaurora	Tech News Website	446207	163
viveraxo	Twitdraw	458279	165
fireflower_	Political Cartoons	452832	195

Table 4: Users with very few followers but high relative influence

Ciertos temas logran atraer una cantidad ingente de atención, por lo que sobresalen sobre el resto en términos de popularidad. A través de su visibilidad, estos temas populares contribuyen en la toma de conciencia colectiva de lo que es

tendencia en la actualidad y, a veces, también puede afectar a la agenda pública de la comunidad. Se trata de los *Trending Topic* o «tema del momento» o «tendencia» (TT), nombre que reciben las palabras o frases más empleadas en un momento concreto en Twitter y que son seleccionados por un algoritmo propietario del propio Twitter. Los diez más relevantes se muestran en la página de inicio, pudiendo el usuario escoger el ámbito geográfico que prefiera -mundial o localizado.

¿Qué factores influyen en que el hecho de que un contenido se convierta en *Trending Topic*?

Asur, Huberman, Szabó y Wang (2011) investigaron las causas que generan un *Trending Topic* sobre los datos analizados de más de 16 millones de *tweet* capturados en un periodo de 40 días en septiembre-octubre de 2010⁷¹. Sus conclusiones más relevantes fueron las siguientes:

- Ley de potencias en la persistencia o duración de los TT. La mayoría de los temas tienen una breve duración, mientras que unos pocos perduran en el tiempo.
- El TT no es resultado de un único autor que lo menciona constantemente, sino que alcanza su posición debido a que muchos autores lo propagan a la vez a través de la red.⁷²
- La naturaleza del contenido es más decisiva en el logro de un TT que las propiedades influyentes (alto grado de actividad y número de seguidores) del usuario que lo inicia.⁷³

⁷¹ Primero capturaban los TT para en los 20 minutos posteriores, recolectar todos los tweets que los mencionaban.

⁷² Esta afirmación coincide con la vertida en el trabajo dedicado a Facebook que admitía en esta red los contagios que se convierten en cascadas globales se dan en eventos que comienzan en un gran número de nodos que ponen en marcha un gran número de cadenas cortas, cada una de las cuales colisiona con rapidez en una sola estructura de gran tamaño. (Sun et al., 2009)

⁷³ Concordante con otros estudios mencionados en los que se ha probado que el número de seguidores no constituye un indicador de influencia.

- La principal forma para propagar información en Twitter es el RT o *retweet*. El número de veces que un tema es retuiteado mantiene una estrecha correlación con la duración temporal del TT.
- Los autores más influyentes, aquellos cuyos contenidos han sido retuiteados en al menos 50 TT, son las fuentes de noticias como CNN o The New York Times. Este hecho pone de manifiesto que más que fuente de noticias, las redes sociales funcionan como filtros y amplificadores de las noticias emitidas por los medios tradicionales⁷⁴.

Tabla 2. Listado de usuarios con contenido más retuiteado extraída del trabajo de Asur et al. (2011)

Author	Retweets	Topics	Retweet-Ratio
vovo_panico	11688	65	179.81
cnnbrk	8444	84	100.52
keshasuja	5110	51	100.19
LadyGonga	4580	54	84.81
BreakingNews	8406	100	84.06
MLB	3866	62	62.35
nytimes	2960	59	50.17
HerbertFromFG	2693	58	46.43
espn	2371	66	35.92
globovision	2668	75	35.57
huffingtonpost	2135	63	33.88
skynewsbreak	1664	52	32
el_pais	1623	52	31.21
stcom	1255	51	24.60
la_patilla	1273	65	19.58
reuters	957	57	16.78
WashingtonPost	929	60	15.48
bbcworld	832	59	14.10
CBSnews	547	56	9.76
TelegraphNews	464	79	5.87
tweetmeme	342	97	3.52
nydailynews	173	51	3.39

Table 1: Top 22 Retweeted Users in at least 50 trending topics each

⁷⁴ Esta conclusión reforzaría, más que la hipótesis de la influencia “two-step-flow” (Katz y Lazarsfeld, 1955) la teoría de Dodds y Watts (2007) según la cual la probabilidad de adoptar una innovación se incrementa cuando un grupo de usuarios, y no sólo uno, repite el mismo mensaje.

Los resultados de sus investigaciones coinciden ampliamente con los presentados por el resto de trabajos mencionados: la popularidad o número de seguidores no es sinónimo de influencia y la naturaleza del contenido es básica para su difusión, siendo aquel de carácter noticiable el más extendido.

“We find that traditional notions of user influence such as the frequency of posting and the number of followers are not the main drivers of trends, as previously thought. Rather, long trends are characterized by the resonating nature of the content, which is found to arise mainly from traditional media sources. We observe that social media behaves as a selective amplifier for the content generated by traditional media, with chains of retweets by many users leading to the observed trends.” (Asur et al., 2011, p.2)

2.5. Política y red

2.5.1. El contagio político

¿Por qué votamos si nuestro voto cuenta poco? ¿Qué nos mueve a hacerlo cuándo intuimos que nuestra participación, por si misma, no es decisiva? ¿Decidimos solos o los demás nos animan? Intentaremos dar respuesta a estas preguntas analizando algunas de las principales teorías conocidas sobre aquello que nos influye y nos lleva a participar en un proceso electoral depositando nuestro voto en las urnas.

“La razón de que votemos, en realidad, tiene mucho que ver con nuestra pertenencia a grupos y con el poder de las redes sociales.” (Christakis y Fowler, 2010, p.187)

Tomamos decisiones calculando todos los costes y beneficios que dicha decisión supone. ¿Cuánto estaríamos dispuestos a pagar de nuestro bolsillo para decidir que ganara uno u otro candidato? ¿Si el hecho de que votemos únicamente es decisivo en caso de empate, y las posibilidades de éste son alrededor de una entre un

millón, qué probabilidad existe de recuperar nuestra inversión? ¿Qué beneficios esperamos obtener a cambio de los costes que nos supone ir a votar, entre los que se incluyen recabar información, analizarla, compararla, decidirse y acudir al colegio electoral?

Anthony Downs⁷⁵ aplicó la ciencia de la “racionalidad” al universo de la política, partiendo del supuesto de que tanto candidatos como votante poseen una visión racional en términos de coste/beneficios, por lo que cada persona decidiría ir a votar tomando en cuenta los costes y los beneficios de su acción: *“Se parte del supuesto de que tanto los votantes como los candidatos poseen una visión racional en términos de costo beneficio, la cual les permite elegir la mejor alternativa que maximice su bienestar.”* (Mantilla y Pinilla, 2008, p.198) Tomando en cuenta la relación costes/beneficios interrogada en el párrafo anterior, *“el análisis racional de la decisión de votar sugiere que ésta equivale más o menos a la decisión de gastarse un dólar en un billete de lotería que nos da una posibilidad entre un millón de ganar un premio de 1.000 dólares.”* (Christakis y Fowler, 2010, p.191)

Por tanto (se obvia, aunque se conoce, el peso del “valor cívico” que puede empujarnos a votar), el análisis racional de ir a votar nos resulta insuficiente. La teoría de la racionalidad pasaba por alto una particularidad esencial: “no votamos solos”. P. Lazarsfeld, N. Berelson y H. Gaudet⁷⁶ investigaron sobre la propagación social de los comportamientos políticos. Una de sus principales conclusiones es que los medios de comunicación no alcanzan directamente a la masa de votantes y que unos “líderes de opinión” actúan como intermediarios entre ellos y amigos y familiares, filtrando la información.

⁷⁵ DOWNS, Anthony. (1957). "An Economic Theory of Political Action in a Democracy". *Journal of Political Economy*, Vol. 65, No. 2 (Apr., 1957), pp. 135-150. The University of Chicago Press.

⁷⁶ LAZARSFELD, Paul F., BERELSON, Bernard y GAUDET, Hazel. (1944) “The people choice’s. How the Voter Makes Up his Mind in the Presidential Campaign”. *Universidad de Columbia*. Nueva York

Posteriormente, R. Huckfeldt y J. Sprague llevaron a cabo un estudio en el que a los encuestados se les pedía que hablaran sobre los amigos que les influían. Comprobaron que las personas con determinada ideología tienden a juntarse y a conversar con individuos similares, actuando de forma consensuada. (Christakis y Fowler, 2010, p.195)

Estos autores habían demostrado que ese cierto grado de polarización, aunque no afectara a la afluencia general a las urnas, sí suponía que si yo decidía apoyar a un candidato podría influir en individuos con ideas afines a las mías y traducirse mi acción-mensaje en votos extra para mi aspirante. Tomando como referencia las conclusiones de estos estudios anteriores, Fowler diseñó un programa informático que simulara lo que sucede en la totalidad de una red cuando una persona decide ir a votar⁷⁷. Sus conclusiones refieren que la afluencia a las urnas es producto de un fenómeno local que se produce en el entorno de marcados sectores de población entre los que se dan pocos grados de separación entre sus miembros: *En algunos casos el voto de un individuo se propagaba como el fuego, desencadenando una cascada de hasta cien personas más votando, aunque de media cada individuo sólo estaba conectado con dos o tres. De media, la decisión de una sola persona de votar motivaba a tres personas más a hacer lo mismo.* (Christakis y Fowler, 2010, p.199)

En “The Link between Social Interaction and Political Choice”, Valdis Krebs investiga, evaluando elecciones presidenciales en EEUU desde el año 1992 hasta el 2000, quiénes pueden ser votantes influyentes, es decir, qué personas pueden influir en otros a la hora de decidirse por una elección. El autor sostiene que el hecho de pertenecer a un grupo homogéneo no supone que todos sus componentes voten a la misma opción electoral; la polarización de grupos, por tanto, no sostiene que todos votemos de igual manera.: *“Common wisdom expects all people in a single network, or cluster to vote for the same candidate. Evaluation of voter selections in the 2000 presidential election found that not to be true. Researchers estimated that only one-*

⁷⁷ FOWLER, James H. (2005) "Turnout in a Small World" Alan Zuckerman, ed. *Social Logic of Politics*. Philadelphia: Temple University Press, 269-287.

third of the voters who chose George W. Bush or Al Gore were in homogeneous networks where all members made the same choice.” (Krebs, 2005)

Sus recomendaciones dirigidas a focalizar una estrategia de contagio político en una red social son las siguientes:

- Una persona puede incrementar el porcentaje de participación anunciando sus planes de ir a votar. Debe hacerse en una comunidad favorable a tu candidato.
- Busca comunidades en las que la opción por el candidato contrario no sea mayoría. Construye relaciones entre los favorables y los indecisos.
- Escanea las comunidades de mayoría contraria y encuentra en ellas personas que respalden a tu candidato. A cada una de ellas, ofrécela un vínculo de soporte.
- Si están integrados en su comunidad y son reconocidos por sus vecinos por su prestigio, plantéate hablar con ellos sobre tu candidato.
- A menos que sean figuras públicas reconocidas, los extraños no influyen en las decisiones de la comunidad. En vez de reclutar desconocidos para llamar a votantes o a las puertas de sus casas, provéete de personas bien conectadas -well-connected- en el entorno en que son valorados.

La investigación anterior tiene mucho que ver con los conceptos “influyentes/líderes de una comunidad” y “vínculos fuertes y débiles”. Se ha referenciado por que manifiesta que la elección del voto es contagiosa.

Así, nuestra decisión de votar influye, y siguiendo la regla de los Tres Grados de Influencia, podría extenderse desde nuestros amigos, hasta los amigos de éstos, y los amigos de los amigos de nuestros amigos. El voto, al fin, es una idea más, un mensaje, una tendencia a la que deberíamos aplicar, para su posterior análisis, los

supuestos incluidos en epígrafes anteriores como la conducta y propiedades de los adaptadores tempranos o las longitudes de las cadenas de difusión.

Continuando con los fenómenos derivados de la influencia que afectan a la decisión final de voto, cabría preguntarnos si el simple hecho de pertenecer (estar afiliado) a una red social (física en este caso) de cualquier índole, sean cuales fueran las condiciones que nos llevaron a participar o convivir con ella, cabría preguntarnos en qué orden se suceden los acontecimientos de contagio político:

- Las personas, con el paso del tiempo, ¿pueden contagiarse de las ideologías políticas de aquellos terceros con los que conviven? o
- ¿Establecemos contactos con aquellos que manifiestan similares opiniones o puntos de vista?

Lazer, Rubineau, Rubineau, Rubineau y Neblo (2009)⁷⁸ encontraron que aunque con el tiempo nuestras actitudes se asemejan a las de aquellos con los que convivimos y conversamos, también es cierto que a la hora de relacionarnos buscamos coincidencias previas. En el caso de la ideología, los individuos cambian sus tendencias políticas hacia las que mantienen sus asociados, y mientras que tendemos a juntarnos con aquellos que comparten nuestras opiniones, esta coincidencia tiene un peso menor a la hora de establecer relaciones que los factores demográficos e institucionales⁷⁹. Por tanto, la ideología política afecta, pero su relevancia a la hora de formar comunidad o establecer relaciones de amistad o de otro tipo, es superada por otra serie de circunstancias que nos rodean. En estas afirmaciones vemos los temas claves que han dominado en la literatura dedicada al comportamiento de las redes sociales:

⁷⁸ Seguimiento de una cohorte de 164 estudiantes de un máster académico en políticas públicas, en dos caminos: su evolución relacionada con la variable “actitud política” y, por otro lado, su evolución de la estructura de su red.

⁷⁹ Esta afirmación fue diferente en Keller y Berry, para quienes las coincidencias demográficas no derivan en una influencia mayor (Keller y Berry, 2003:31): “*There is a certain demographic “center of gravity” to Influentials – college educated, midlife,... in positions of responsibility in the workplace- but they are not a uniform group by conventional measures.*”

- La homofilia, como la tendencia de individuos similares a compartir lazos o conexiones similares.
- La influencia, como la tendencia de las personas que comparten algo a convertirse en objetos bastante iguales.

Siguiendo a Lazer et al. (2009), la homofilia no viene impuesta por la estructura social ni la oportunidad, es el resultado de factores endógenos que facilitan la atracción: motivo informacional, motivo preferencial y motivo de reafirmación. En este caso, la relación que se establece entre dos personas con orientación política similar es más probable que se convierta en una relación duradera. Además, en un entorno dominado (los individuos son mayoría numérica) por una ideología, los contrarios tienden a la no participación en las actividades que se sucedan.

Respecto a la influencia, nuestras redes nos afectan continuamente en este sentido. De hecho, nos volvemos más afines con aquellas personas con las que mantenemos una relación continuada. La orientación política no sobresale como factor contagioso⁸⁰, de hecho, en ausencia de conversaciones sobre este tema, la influencia ideológica (aún conocida la preferencia de nuestro nexo) es limitada o inexistente. Sin embargo, las tendencias políticas de una persona que nos resulta atractiva (más allá de la atracción física) o con la que mantenemos una estrecha relación, produce en nosotros un impacto mucho mayor que idénticas opiniones vertidas por personas con las que la relación establecida es circunstancial (por ejemplo, compañeros de trabajo). (Lazer et al., 2009, pp.5-6).

Este hallazgo afectaría a la forma en que el contagio, en forma de influencia, se sucedería en redes sociales en Internet, en las que un cierto número de las relaciones que se establecen entre los individuos enlazados corresponden a relaciones del tipo “circunstancial” o “superficial” o de *small talks*. Lazer et al. (2009, p.21) remarcan que dado que la mayoría de las conversaciones sobre asuntos políticos se

⁸⁰ Para evitar tensiones en una relación dada somos capaces de cambiar nuestras opiniones y/comentarios acerca de un candidato.

suceden en este tipo de conexiones y no en las relaciones estrechas (lazos fuertes), y dado estas últimas provocan en nosotros una mayor conversión o transformación ideológica, la influencia política es más una consecuencia de los sentimientos de afecto que de la simple transmisión o circulación de opiniones a través de cualquier sujeto de la red. (Lazer et al., 2009, p.21)

Frente a estas últimas afirmaciones de su investigación, es necesario poner de manifiesto que si bien Internet y sus Redes Sociales pueden denotar cierto grado de homofilia resultando en grupos polarizados⁸¹, el seguimiento o el simple nexo con individuos reconocidos en ellas como “influyentes”, aún no manteniendo un individuo con ellos una relación social basada en el afecto o preferencia emocional, podría considerarse una anexión de “lazo fuerte” con una destacable capacidad de conversión debida a la alta capacidad de influencia derivada de su renombre o posición social o profesional destacada y reconocida en el entorno en el que se desenvuelve. Además, si bien los “lazos débiles” son especialmente aplicables a la difusión de innovaciones⁸², ¿no podrían los partidos y candidaturas de reciente aparición, por su carácter de innovación, propagarse al menos, cuando no ser adaptadas, por un grupo determinado de individuos propenso a practicar en la novedad?

¿Quiénes son los influyentes en una sociedad-red?

Generalmente la influencia se asocia a una persona con dinero y popularidad, pero las personas reconocidas como tal (Roosevelt, Kennedy, Oprah, etc.)⁸³ no suponen la mayor influencia en el conjunto de la sociedad. Si bien sus posiciones destacadas favorecen cierto grado de influencia, no ellas ni sólo ellas afectan al comportamiento de los demás. El contagio real se resuelve mayormente entorno a

⁸¹ Más adelante dedicamos un epígrafe a la polarización de grupos en la Red.

⁸² Granovetter (1973) y Cha et al. (2010).

⁸³ Referidas a los Estados Unidos de América.

comunidades en las que alguno o varios de sus componentes adquieren la figura de “influyente”:

“Decades of research on change or trends in the society have convinced us, however, that there is a major source of influence in America that lies with a larger group of Americans.” (Keller y Berry, 2010, p.30)

Durante décadas, según las investigaciones conducidas por la firma de consultoría e investigación de mercado RoperASW, que monitorizó las tendencias de mercado durante 30 años para muchas de las compañías americanas que aparecen publicadas en el ranking de Fortune 500, los influyentes han marcado y dirigido el camino del desarrollo social en EEUU, desde el resurgimiento de la autonomía para manejar sus decisiones de consumo hasta el escepticismo y la falta de confianza en los anuncios de marketing y las propagandas políticas. Educados, de mediana edad, y política y socialmente activos, son personas que utilizan más canales de información y de comunicación que la media, entre los que ya empezaban a despuntar Internet y el correo electrónico junto con los medios de comunicación tradicionales: *“Influentials are more likely to use not only mail bu e-mail and computers to communicate tan the public as a whole is to send a letter (see Figure 3-4).”* (Keller y Berry, 2010, p.139)

En el año 2006, David Nickerson realizó un estudio experimental sobre afluencia a las urnas demostrando que animar a una persona a votar era suficiente para influir en otra persona que no había recibido el mensaje inicial personalmente⁸⁴. ¿Y si la cadena de influencia no se detiene ahí y pasa a otras? Todo dependería de la forma y características de nuestra red (transitividad, número de conexiones, tipo de vínculos, corrupción del mensaje, etc.). Fowler, aunque asume que cada persona tiene a más de dos amigos cercanos en los que poder influir, se pregunta si el rebote

⁸⁴ “Is voting contagious? Evidence from Two Field Experiments” *American political Science Review* 102 (2008: pp. 45-57).

del mensaje en grupos cerrados nos impediría influir en personas alejadas de nosotros y que el mensaje se deteriore antes de llegar a esos individuos alejados. (Christakis y Fowler, 2010, p.202). El autor no ha tenido en cuenta la importancia de la fuerza de los vínculos débiles ni las especiales características de Internet en la transmisión de tendencias.

Consideramos que si aceptamos que la intención de voto es un acto de naturaleza contagiosa y las redes sociales en Internet son capaces de manifestar características propias –la influencia de nodos no centrales, la longitud y forma de la cadena de difusión, multiplicación de conexiones, proliferación de vínculos débiles– que difieren de las teorizadas en las redes tradicionales, en este canal, el contagio político puede adquirir valores numéricos y comportamientos que sorprenderían por su cantidad y procedimientos únicos.

2.5.2. Internet y la polarización de grupos

Las nuevas tecnologías pueden conectar a personas separadas geográficamente y ayudarles a encontrar recursos e informaciones útiles, lo que, de hecho, serviría para reducir las diferencias y formar comunidades. Por el contrario, las mismas habilidades podrían, paradójicamente, fragmentar la interacción y dividir los grupos. En Internet cada día es más fácil para un individuo fijar la atención en lo que le interesa y eliminar o simplemente no aceptar a contactos poco deseados. En el futuro, las preferencias individuales y las capacidades que aporta la Red para desarrollarlas, serán determinantes en la definición de los límites de las comunidades. (Alstytne y Brynjolfsson, 2004, p.1)

Es conocido el positivismo que Internet supone para el activismo y la comunicación, que suponen que realmente las nuevas tecnologías favorecen el activismo y, por tanto, cooperarán en la generación de nuevas formas de gobierno que superen las tradicionales democracias de hoy e incluso, algunos métodos

utilizados por regímenes autoritarios. En este supuesto, queda aún por reflexionar si el acceso global a la información y el permanecer “conectados” genera auténticas comunidades o individuos aislados, si las características sociodemográficas en su acceso provoca diferencias entre los individuos y las sociedades, y si fomenta el fenómeno de la polarización de grupos por el que los miembros de un grupo deliberativo se mueven hacia un punto más extremo en la dirección indicada por sus tendencias ideológicas.

Respecto a la primera cuestión, existen corrientes de opinión que afirman que el acceso a Internet causa, finalmente, aislamiento en quienes la utilizan. Según el informe de *Pew Internet* dedicado a este tema y publicado en noviembre de 2009, “*Social Isolation and New Technology*” los norteamericanos no están tan aislados como se había reportado en estudios previos realizados en diversas universidades, y el móvil y la presencia en redes sociales se asocia con una mayor participación en actividades sociales que requieren contacto físico. Internet influye positivamente en una mayor participación en las comunidades locales y su uso anima a continuar la vida social en lugares públicos como parques o restaurantes, y, contrariamente a lo que se daba por hecho, se utiliza más para fortalecer los contactos cercanos que para mantener relaciones a larga distancia. (Hampton, Sessions, Her y Rainie, 2009)

La “brecha digital” fue uno de los primeros fenómenos de diferenciación social que se produjo con la llegada de Internet. La primera brecha se conoce como la que separa a aquellos individuos o sociedades que pueden conectarse de los que no, debido en su mayor parte al contexto socioeconómico poco desarrollado en el que los primeros se desenvuelven. También se habla de brecha digital para referirse al vacío que se contempla en las personas de edad avanzada, con escasos conocimientos en las nuevas tecnologías.

Desde entonces, ha habido mucha discusión sobre la cuestión de raza y clase en el contexto de Internet, y los medios sociales en particular. Mientras que la

mayoría de los primeros estudios se centraron en el acceso (“The Digital Divide”⁸⁵), los investigadores han ido cambiando su enfoque a la diferenciación en el ámbito de aplicación y el uso de la Internet para fines distintos (DiMaggio et al., 2004)⁸⁶. Recientemente, algunas investigaciones han sugerido que la pertenencia a Redes Sociales en Internet es un proceso cada vez más selectivo (Boyd, 2009)⁸⁷, y que el uso de Internet se está convirtiendo en un medio de comunicación social diferenciado por el nivel socioeconómico (Hargittai y Walejko, 2008)⁸⁸. (Chang et al., 2011, p.7)

Otra de las paradojas que ha arrojado el desarrollo de las nuevas tecnologías, y más concretamente Internet, es si su uso facilita la creación de grupos polarizados, es decir, la tendencia por la que las personas se volverían más extremistas después de haber hablado o contactado con otros que comparten su misma opinión.

La polarización de grupos vendría a ser una consecuencia directa de la homofilia⁸⁹ o tendencia de las personas a relacionarse con personas que se parecen a ellas respecto a diferentes atributos como creencias, clase social, educación, edad, aficiones, intereses comunes, etc. En la polarización de grupos se tiende a fortalecer las inclinaciones promedio de sus participantes y aplicado a la deliberación en la esfera pública de una democracia: *“En resumen, la polarización de grupos significa que los miembros de un grupo deliberativo se mueven previsiblemente hacia un punto más extremo en la dirección indicada por sus tendencias pre-deliberativas.”* (Sunstein, 2001, p.55)

⁸⁵ NORRIS, Pippa (2000). "The Worldwide Digital Divide: Information Poverty, the Internet and Development" *Paper for the Annual Meeting of the Political Studies Association of the UK, London School of Economics and Political Science, 10-13th April 2000*. Roundtable on The Future Role of New Media in Elections, Wednesday 12th April 10.45-12.15.

⁸⁶ DIMAGGIO, Paul, HARGITTAI, Eszter, CELESTE, Coral y SHAFER, Steven. (2004). “From unequal access to differentiated use: A literature review and agenda for research on digital inequality”. *New York, NY: Russell Sage Foundation*. 355–400.

⁸⁷ BOYD, Danah. (2009). "The Not-So-Hidden Politics of Class Online." *Personal Democracy Forum, New York, June 30*.

⁸⁸ HARGITTAI, Eszter, y WALEJKO, GINA (2008). “The participation divide: Content creation and sharing in the digital age”. *Information Communication and Society 11(2):239*

⁸⁹ Véase en este capítulo Lazer et al. (1955)

Dado que Internet facilita el contacto entre personas con intereses similares que de otra forma no podrían establecer contactos, al menos, no fácilmente, se ha extendido la idea de que la Red favorece el efecto de la polarización de grupos. En el campo de las política de comunicación y deliberación de masas, el efecto de la polarización de grupos se ve favorecido por la facilidad para atender las voces que concuerdan con nuestra forma de pensar: *“Con una mayor especialización, los individuos están cada vez en mejores condiciones de evitar los periódicos y las revistas de interés general, y de escoger opciones que reflejen sus propias predisposiciones. Internet está haciendo posible que la gente diseñe sus propios –y altamente personalizados– paquetes de comunicación, filtrando cuestiones urticantes y aquellas voces no favorecidas.”* (Sunstein, 2001, p.64) Para Sunstein, el problema central de tal polarización se encuentra en que es probable que se produzca una fragmentación social y se difundan errores cuando los individuos que piensan de manera similar se encuentren aislados y se muevan en direcciones extremas.

Sunstein, continúa su visión un tanto negativa de Internet en su libro *Republic.com: Internet, Democracia y Libertad*, afirmando que polariza enormemente el discurso político porque permite a los individuos elegir la información que quieren e interactuar preferiblemente con otros que mantienen su misma opinión. La participación anónima, que según Sunstein es típica en la vida online, despersonaliza al individuo siendo más susceptible de adherirse a un grupo y aceptar sus normas.

Cuatro años más tarde, el investigador en ciencia política Peter Muhlberger realizó un estudio encuestando a mil doscientos adultos en el área de la ciudad de Pittsburgh. Presentó sus resultados y discusiones a la teoría de Sunstein en un paper presentado en el 2004 *International Communication Association Annual Meeting en Nueva Orleans*. Con el título “Polarization of Political Attitudes and Values on the Internet”, Muhlberger revisa diferentes teorías acerca de este fenómeno, y junto con los datos recogidos en su trabajo de campo en Pittsburgh, concluye afirmando que en

vez de dividir la sociedad, Internet puede servir como una influencia moderadora (Muhlberger, 2004, pp.1-28):

- Los puntos de vista sobre el efecto de Internet en la intensificación del fenómeno de polarización de grupos han sido alarmistas.
- Estadísticamente, las diferencias entre polarizaciones en grupos online y offline apenas se aprecian.
- En Internet, la tendencia es hacia una menor polarización y una mayor agrupación, lo que sugiere que los usuarios de la Red se están moviendo hacia el centro de las actitudes y valores políticos.

La polarización ideológica de los medios de comunicación tradicionales es una condición constantemente anunciada y evidenciada. Algunos trabajos han mostrado como los consumidores de noticias tienen una preferencia significativa por los medios de comunicación afines (liked-minded)⁹⁰, y que utilizan las herramientas sociales de Internet para buscar aquellos con similares puntos de vista⁹¹.

¿Se replica dicha polarización en Redes Sociales como Twitter?

Para comprender las preferencias políticas de las audiencias de diferentes medios de comunicación en Twitter, Golbeck y Hansen (2010) aplicaron un método para computar dichas preferencias utilizando las puntuaciones (1.0 para liberal/ 0.0 para conservador) que publica la asociación *Americans for Democratic Action* (ADA) sobre los congresistas⁹². Una vez capturados los seguidores de cada congresista y computados (*P Score*) con un valor en función de la ideología del congresista o congresistas a los que siguen en Twitter, se establecen correlaciones con la cuentas en Twitter de los principales medios de comunicación (indicados por

⁹⁰ GELMAN, A. "Red state, blue state, rich state, poor state: why Americans vote the way they do". *Princeton Univ Press, Princeton NJ, 2008*.

⁹¹ Garret, K. "Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate". *Journal of Communications, 59, 4 (December 2009), 676-699*.

⁹² "Americans for Democratic Action. Annual Voting Records, 2009". Accesible en <http://www.adaction.org/>. (Consultado el 20/10/ 2011)

Gentzkow y Shapiro⁹³). Los autores refutaron en esta red el comportamiento polarizado y/o de tendencia a la homofilia ideológica, de hecho, entre todos los seguidores de políticos encontrados, un 62,5% sigue a aquellos que comparten su ideología contra un reducido 5% que apoyan opciones opuestas, siendo el resto un grupo con mezcla de apoyos. Su conclusion final es que: *“Our results show that the political preferences of media outlets’ audiences reflect the liberal/conservative leanings of the media organizations as presented in prior literature.”* (Golbeck y Hansen, 2010, p.4)

“Who users choose to follow on Twitter affects which tweets they see.” (Golbeck y Hansen, 2010, p.3)

Tabla 3. Porcentaje de audiencia de medios de comunicación en Twitter, extraída del trabajo de Golbeck y Hansen (2010)

Media Outlet	Twitter User ID	Total Followers	Avg. % of Total Followers sampled	Avg. P Score
Fox News	foxnews	266,121	7.33%	0.26
The Drudge Report	Drudge_Report	102,981	20.52%	0.27
Washington Times	washtimes	13,545	33.75%	0.29
Wall Street Journal	WSJ	392,332	5.75%	0.41
US News & World Report	usnews	7,836	28.60%	0.49
The L.A. Times	latimes	72,296	9.91%	0.51
USA Today	USATODAY	63,714	9.90%	0.51
Good Morning America	gma	1,698,875	0.72%	0.51
The News Hour	NewsHour	54,787	12.79%	0.52
CBS News	CBSNews	1,578,599	0.94%	0.53
Newsweek	Newsweek	1,256,536	1.25%	0.54
Washington Post	washingtonpost	154,400	9.51%	0.54
The Today Show	todayshow	629,088	1.89%	0.55
The Early Show	theearlyshow	12,628	11.16%	0.55
Time Magazine	TIME	2,134,411	0.99%	0.57
ABC World News	abcworldnews	12,973	14.93%	0.57
CNN Breaking News	Cnnbrk	3,314,716	1.40%	0.58
NBC Nightly News	nbcnightlynews	27,137	16.85%	0.58
The New York Times	nytimes	2,553,291	1.40%	0.60
Morning Edition (NPR)	MorningEdition	6,146	17.22%	0.66

Table 1. The average audience P Scores for the twenty popular media outlets studied in [6]. Averages are computed over 10 randomly drawn samples using the sampling method described above. Results are sorted from most conservative to most liberal.

⁹³ Gentzkow, M. and Shapiro, J. “What drives media slant? Evidence from US daily newspapers”. *Econometrica*, 78, 1 (2010), 35–71.

En el trabajo de Golbeck y Hansen (2010), aunque destinado a conocer las preferencias ideológicas de las audiencias de los medios de comunicación (y de su tipología deducir también la tendencia política de dichos medios), se ha puesto en evidencia la homofilia política que se da entre los usuarios de Twitter en lo que respecta a los políticos que siguen en esta red. Pero en Twitter no sólo se cuenta con los seguidores o “followers” para determinar el carácter de una conexión, sino que gran parte de su conversación se lleva a cabo a través de dos de sus herramientas internas más populares: las menciones y los *retweets*⁹⁴. Conover et al. (2011), de la Universidad de Indiana, llevaron a cabo un análisis enfocado a profundizar en la polarización política en Twitter.

Recolectando más de 250.000 *tweets* publicados en las semana anteriores a las elecciones legislativas de 2010 en EEUU, e identificando la tendencia del contenido político de los mismos a través de la catalogación manual de los textos que se utilizan como *hashtag*, demostraron que, mientras que la red de *retweets* aparece como una red altamente partidista con escasas conexiones entre usuarios de “izquierda y derecha”, la red de menciones aparece como un grupo homogéneo en el que individuos con visiones políticas diferentes conversan entre sí en un grado mucho más grande que el que se obtiene en la red de *retweets*. (Conover et al., 2011)

⁹⁴ Los *retweets* funcionan como una forma de respaldo hacia una opción determinada, permitiendo a los individuos retransmitir contenidos generados por otros usuarios, aumentando así la visibilidad del contenido afín. Las menciones cumplen una función diferente, permitiendo dirigirse directamente a un usuario específico, o, en menor medida, referirse a un individuo en tercera persona.

Ilustración 9. Visualización polarizaciones subred “retweet” y subred “mention” de Conover et al. (2011)

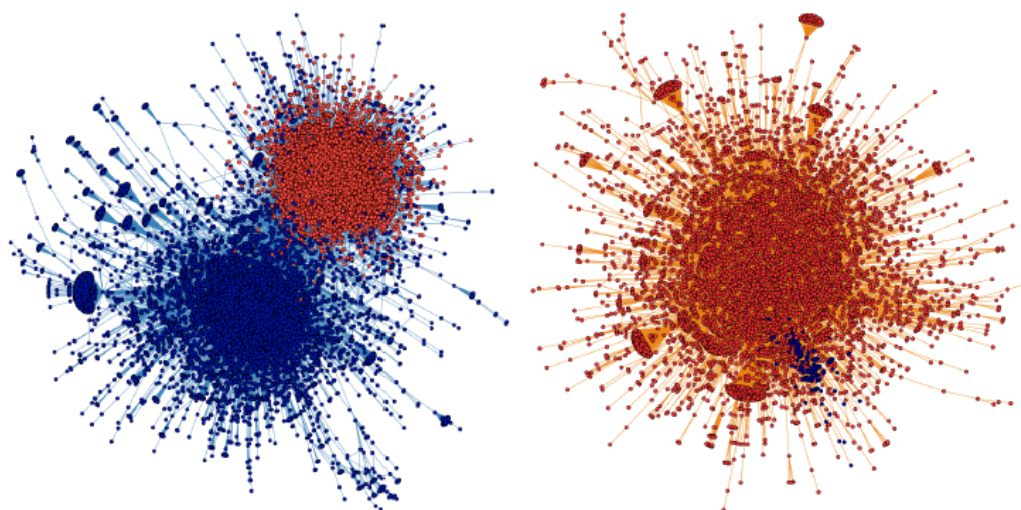


Figure 1: The political retweet (left) and mention (right) networks, laid out using a force-directed algorithm. Node colors reflect cluster assignments (see § 3.1). Community structure is evident in the retweet network, but less so in the mention network. We show in § 3.3 that in the retweet network, the red cluster A is made of 93% right-leaning users, while the blue cluster B is made of 80% left-leaning users.

Respecto a la polarización de grupos y su comportamiento más o menos extremo, y coincidiendo con las teorías de Sunstein, observaron que a pesar de que los usuarios con posiciones ideológicas contrarias interactúan entre ellos, es muy escasa la información que comparten desde el otro lado de la brecha con otros miembros de su comunidad. Esto es posiblemente debido a que las presiones sociales que conducen a la polarización de grupo son demasiado fuertes para la mayoría de los usuarios como para superarlas. Además, muchos de los mensajes publicados contienen expresiones más extremas que las que pueden esperarse en una relación cara a cara. Cristakis y Fowler (2010, p.208) admiten que existe “la opinión generalizada de que en los últimos años se ha intensificado en Estados Unidos la polarización entre demócratas y republicanos”.

El estudio de la blogosfera americana durante las elecciones de 2004 en Estados Unidos puso de manifiesto que la red social online de blogs era profundamente homófila y está muy polarizada. Adamic y Glance (2005) analizaron los post publicados en los 40 principales blogs políticos en el plazo de dos meses

anteriores a las elecciones presidenciales de los Estados Unidos de América⁹⁵ en 2004, con el objeto de estudiar con qué frecuencia se referían el uno al otro y para cuantificar la coincidencia en los temas que abordaron, tanto dentro de la comunidad liberal como de la comunidad conservadora, y también entre comunidades. También capturaron una instantánea de un solo día de más de 1.000 blogs políticos, de la que obtuvieron una fotografía de los blogroll (la lista de enlaces a otros blogs) de dichos blogs. Después de un análisis detallado de las conversaciones/relaciones entre blogs, junto con los nombres de políticos y medios de comunicación que mencionaban en sus mensajes, se encontraron con dos comunidades diferenciadas, liberal y conservadora, siendo el efecto de refuerzo mayor todavía en la derecha política y no encontrando, sin embargo, que en esta comunidad una monotonía mayor que en la liberal a la hora de mencionar o referenciar noticias o temas afines: *“An interesting pattern that emerged was that conservative bloggers were more likely to link to other blogs: primarily other conservative blogs, but also some liberal ones. But while the conservative blogosphere was more densely linked, we did not detect a greater uniformity in the news and topics discussed by conservatives.”* (Adamic y Glance, 2005, p.14)

⁹⁵Blogs listados en la categoría de blogs políticos en varios directorios de blogs en línea, incluyendo eTalkingHead, BlogCatalog, Línea de Campaña, y Blogarama

Ilustración 10. Visualización de la polarización de la blogosfera americana en las elecciones de 2004, de Adamic y Glance (2005)

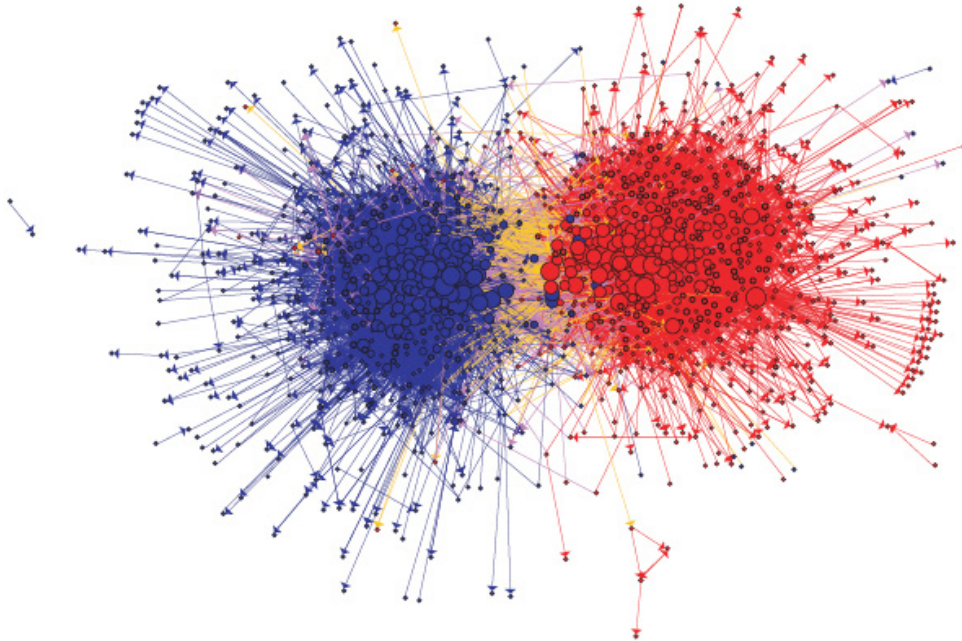


Figure 1: Community structure of political blogs (expanded set), shown using utilizing a GEM layout [11] in the GUESS[3] visualization and analysis tool. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

2.6. Conclusiones

Internet, en un principio, y las redes sociales más tarde, han dejado patente que los medios o canales en los que la política se venía desarrollando ni son los mismos ni siguen las mismas pautas. Lo que sabíamos o intuíamos acerca de cómo implementar una eficaz comunicación política en campaña electoral ya no sirve, porque en Internet los patrones de comportamiento y difusión parecen no obedecer a las premisas, estudiadas y probadas, que teníamos adquiridas. Las redes sociales físicas han existido desde siempre y su análisis ha sido objeto de estudio por que forman parte de la vida misma, son causa del mundo tal y cómo lo conocemos, pero en Internet, sus mecanismos podrían no atender a las reglas hasta hoy conocidas.

El análisis de la teoría de redes sociales, desempeñado por sociólogos, biólogos y antropólogos, entre otros, ha intentado comprender el funcionamiento de una red social, buscando patrones de comportamiento en sus actores –nodos- y en las relaciones -vínculos- que se establecen entre ellos. Son muchos los estudios dedicados al diseño de su morfología y a la definición de sus características y propiedades, tratando de identificar los rasgos fundamentales que influyen en el papel que juega cada actor en la red y los patrones de difusión y contagio de todo aquello que circula por la misma.

Entre todas las conclusiones revisadas en diversos estudios, hemos fijado la atención en los aspectos que guardan una mayor relación con el objeto de nuestro trabajo, ya que nos ayudarían a entender cómo se transmite una idea o mensaje en una red social e intentar, posteriormente, de buscar similitudes o diferencias con las redes sociales que se configuran en Internet: el peso de la centralidad, determinación del grado de influencia, la fuerza de los vínculos débiles, la capacidad de contagio entre individuos y la amplitud de la cadena de difusión.

Las teorías sobre el poder de la centralidad en la estructura de una red social son de gran importancia ya que pueden determinar la capacidad que un actor central adquiere para contagiar lo que transmite, y sustentar la idea de que aquellos nodos bien posicionados -centrales- demuestran cierta supremacía sobre las ideas o comunicación del resto y que, por tanto, alcanzar los objetivos que se planteen -propios o de terceros- les resulta mucho más fácil que al resto. Los estudios “clásicos” realizados en entornos físicos concluyeron que la centralidad, un atributo estructural de los nodos de una red que se relaciona con una mayor posición central (y menos periférica), influye en la percepción de liderazgo, la rapidez y eficiencia de una red en resolver sus problemas y la satisfacción de sus miembros. Sin embargo, cuándo se investiga el peso de la centralidad en Internet, se advierte que la posición central de un nodo deja de ser predominante en la canalización, difusión y contagio de los mensajes, siendo procesos de carácter más igualitario y democrático.

El grado de influencia queda establecido en la “Regla de los tres grados de influencia” indicado por Christakis y Fowler (2010), según la cuál no podemos influir más allá de “los amigos de los amigos de nuestros amigos” debido en parte, a que la información se va deteriorando a cada paso y a que los vínculos menos estables o sin conexión directa se rompen. Pero esta regla podría no darse en las redes sociales en Internet ya que los principales argumentos que la sustentan no se cumplen: la fidelidad de la información se mantendría por la inmutabilidad que se desprende de su carácter electrónico y los vínculos tenderían a ser más estables al no quedar sujetos a las condiciones particulares espaciales o temporales de sus nodos.

Los primeros análisis de redes sociales argumentaban que los vínculos fuertes entre sus nodos eran determinantes en la eficacia de la transmisión de mensajes en ella. Posteriormente llegaron nuevas teorías que hablaban de “la fuerza de los vínculos débiles” y cómo cualquier cosa que pueda ser difundida podrá llegar a un gran número de personas y atravesar una gran distancia social cuando se experimentan vínculos débiles antes que fuertes ya que en grupos cerrados/próximos muchos oirán el rumor en varias ocasiones y puesto que la motivación por extender un rumor se debilita cada vez que éste se cuenta, los puentes de conexión no serán cruzados. Dado que los vínculos débiles son indispensables para las oportunidades individuales y para su integración en la comunidad y que los vínculos fuertes llevan a una fragmentación total, y que los vínculos fuertes ayudan pero los débiles son más útiles porque aumentan el número total de conexiones, las redes sociales en Internet se muestran óptimas para el establecimiento de multiplicidad de vínculos débiles, mostrarían una capacidad mayor para la amplitud del alcance en la difusión de un mensaje, idea o tendencia. Además, se ha demostrado que si bien se consideraba que los vínculos fuertes eran más confiables y creíbles en el momento de adoptar un nuevo comportamiento, en este nuevo canal los pesos de vínculos fuertes y de los vínculos débiles se igualan en valor.

En las redes, el contagio es un factor del mayor interés. Ha quedado validado que las ideas, los productos, los mensajes y las conductas se extienden entre nosotros

igual que los virus. Un “meme” es una unidad de información cultural que puede ser transmitida de una mente a otra, de una persona a otra, de una generación a otra; un modelo o patrón de información retenido en una memoria que es capaz de ser copiado en otra memoria individual. Diversas teorías relacionadas con la adopción de conductas y la difusión de innovaciones, han manifestado que los medios de comunicación sólo son efectivos en el punto de partida de la innovación o noticia siendo las personas las que influyen en la adopción o rechazo por otros, que los vínculos débiles juegan un papel crucial en la difusión de innovaciones, que los nodos definidos como influyentes no son más apropiados para conseguirlo que la media de los nodos, y que, si bien los nodos centrales favorecen la adopción de nuevas conductas, los adaptadores tempranos *-early adopters-* considerados marginales por su posición en la red, resultan imprescindibles en la transmisión de conductas o innovaciones difíciles. Se ha constatado que en Internet ha surgido una nueva clase de influyentes, los “super-influyentes”, cuyas opiniones son admitidas por otros y que responden a la definición de *early adopters*. Si el contagio es posible y la popularidad no siempre es sinónimo de superior influencia y los lazos débiles son una pieza clave en el proceso, las redes sociales configuradas en los servicios que ofrece Internet, contienen los ingredientes óptimos para mejorar un proceso de difusión: alto número de conexiones entre individuos muy diversos, multitud de nodos centrales y marginales, y proliferación de vínculos débiles.

La longitud y forma de la cadena de difusión se ha estudiado en los análisis de las teorías de redes intentando comprender de qué depende, o que influye, en la producción de una difusión en cascada. En ellas se ha visto reforzado el papel de la centralidad como desencadenante así como el protagonismo de ciertos individuos considerados prominentes o influyentes. Si bien Internet podría caracterizarse por poseer unos nodos fuertes o *hubs* –elevado número de conexiones– que son los que la dan forma y la estabilizan, diferentes estudios han concluido que la presencia de individuos altamente conectados es de escasa relevancia en la producción de este fenómeno, y que aunque la centralidad y los lazos fuertes son determinantes en la transmisión y adopción de una innovación, la difusión boca a boca en la red Internet

se revela como bastante descentralizada, y que el modelo de difusión de información en cascada en sus redes sociales no se produce en largas y ramificadas cadenas sino que obedece a un patrón de difusión caracterizado por una larga escala de colisiones de cadenas cortas.

En la red social Facebook algunos autores han determinado que la influencia existe y que ésta funciona igual que en el mundo real. Sin embargo, otros estudios afirman que el modelo de difusión en Facebook difiere de los modelos clásicos en lo que se refiere al peso de los influyentes y a la longitud de las cadenas “infectadas”. Mientras que tradicionalmente se admitía que un evento comienza con un reducido número de nodos iniciadores que son capaces de afectar a los vecinos más vulnerables, en Facebook los contagios que se convierten en cascadas globales se dan en eventos que comienzan en un gran número de nodos que ponen en marcha un gran número de cadenas cortas, cada una de las cuales colisiona con rapidez en una sola estructura de gran tamaño. Por tanto, no se puede identificar a los iniciadores como lo más importante en un proceso de difusión cultural, ya que en redes sociales como Facebook una gran cantidad de individuos tienen la posibilidad de desencadenar una cascada global de gran tamaño. Además, aunque ahora todos podemos influir en los demás, ha surgido la aparición de una nueva clase de influyentes a los que denominamos “super-influyentes”. Individuos que gracias a los social media y a las herramientas que proveen para crear y compartir juicios, superan el grado medio de la influencia difundiendo su opinión más lejos y más ampliamente.

James Fowler, un politólogo de la Universidad de California, San Diego, dirigió la investigación “A 61-million- person experiment in social influence and political mobilization”, concluyendo que hubo un tercio de millón de personas que realmente se presentaron en las urnas que de otro modo no tendrían si el mensaje de intención de voto de sus amigos en Facebook, no les hubiera llegado: *“The results show that the messages directly influenced political self-expression, information seeking and realworld voting behaviour of millions of people. Furthermore, the messages not only influenced the users who received them but also the users’ friends,*

and friends of friends. The effect of social transmission on real-world voting was greater than the direct effect of the messages themselves, and nearly all the transmission occurred between ‘close friends’ who were more likely to have a face-to-face relationship.” (Bond et al., 2012).

La muestra de 61 millones de personas fue dividida en tres categorías. El primer grupo, que contaba con unos 60 millones de usuarios, recibía un mensaje en su perfil de Facebook que decía: "Hoy es el día de las elecciones", acompañado de una figura gráfica con bandas de colores rojo, blanco y azul, en la que rezaba "Vote" y ofrecía, además, ayuda para encontrar un lugar de votación local⁹⁶. Es importante destacar que también se incluyen los rostros de hasta seis amigos que también han optado por hacer clic en la opción "Vote". El segundo grupo, con 600.000 personas, recibe el mismo mensaje, pero sin ninguna información relativa al resto de amigos del emisor. El tercer grupo constaba de 600.000 personas que recibieron ningún tipo de mensaje.

Siguiendo estado por estado los registros electorales para determinar quién había mostrado su intención de depositar su voto, los investigadores encontraron que el primer grupo votó a un ritmo ligeramente más alto que los miembros de los otros dos grupos, lo suficiente como para traducirse en 60.000 votos adicionales.

El resto del aumento se atribuye a contagio social. Las personas que habían visto los mensajes con las caras de sus amigos también eran más propensos a hacer clic en botón "Voté". Esta acción, se trasladaba a su vez, a sus amigos. Y esos amigos también votaron a tasas más altas, traducándose en otros 280.000 votos adicionales.

La red Twitter ha demostrado es una capacidad asombrosa para difundir información entre sus millones de usuarios. Las herramientas de que surte para

⁹⁶ Los mensajes de Facebook de voto no son partidistas, y los autores indican que el efecto de mayor votación fue el mismo en un subconjunto de los votantes autodefinidos como conservadores que en los autodefinidos como liberales.

facilitarla son los *retweet* (RT), los *hashtag* (#) y derivados de estos, los *Trending Topics* o temas de máxima actualidad y actividad en la red. En un análisis realizado sobre la transmisión de mensajes en este canal, se manifestó que se cumple escasamente la “ley de potencias” en la distribución de seguidores, que existe un grado bajo de reciprocidad en las relaciones, que la mayoría de los temas se corresponden con titulares de noticia o similares, que el concepto de popularidad se obtiene tanto por número de seguidores como por la popularidad de un *tweet*, y que tras el primer *retweet*, un mensaje es re-difundido por otros a sus seguidores instantáneamente. En Twitter aunque los usuarios con más seguidores tienen más posibilidad de obtener *retweets* -redifusión- de sus mensajes (por el número de seguidores en sí, obviamente), algunos individuos son capaces de conseguir muchos *retweets* con un bajo número de seguidores debido probablemente a que su contenido se considera interesante y contagioso.

Es imprescindible remarcar que, en lo que respecta al comportamiento de las Redes Sociales en Internet comparado con el comportamiento de las redes sociales físicas, las diferencias en las conclusiones de diversos estudios (ya papables en el análisis de redes físicas) dedicados al estudio de dicho comportamiento, difieren en sus conclusiones, y a más, no se encuentra siempre paridad en los resultados concluyentes entre los mismas (la mayor digresión se halla entorno al concepto de la Ley de Potencias, que de existir, pudiera no ser la única condición significativa ni decisoria en aras de conseguir que una idea cualquiera se propague por la red, y a el peso de los vínculos fuertes versus vínculos débiles). No obstante, si coinciden en aspectos que son sin duda del mayor interés:

- Más que el/los usuario/s influyente/s se trata de un número considerable de nodos que inician un proceso de difusión que concluye en una cascada global.
- La popularidad o número de seguidores así como el volumen de actividad de un usuario, no es sinónimo de influencia.
- La naturaleza del contenido es básica para su difusión, siendo el de carácter noticiable el más propenso a conseguirlo.

- Los influyentes no siempre consiguen que su contenido sea retwiteado. Por el contrario, usuarios anónimos alcanzan un *Trending Topic* (TT), el cual es resultado de que una cantidad suficiente de usuarios lo propaguen a la vez.
- Los lazos débiles y los *early adopter* son de vital importancia en el proceso de difusión de innovaciones.
- La teoría “*two-step-flow*” parece quedar relegada en estas redes, en las que los “influyentes” no son únicamente los responsables de la difusión desde los medios, sino que, una vez más, los escasos vínculos (lazos débiles) posibilitan las interacciones entre comunidades (individuos densamente conectados).

En política también se da el contagio. El análisis racional de ir a votar resultó insuficiente porque la teoría de la racionalidad pasaba por alto una particularidad esencial: “no votamos solos”, el voto de un individuo puede propagarse como el fuego desencadenando una cascada. Así, nuestra decisión de votar influye, y siguiendo la regla de los Tres Grados de Influencia, podría extenderse desde nuestros amigos, hasta los amigos de éstos, y a los amigos de los amigos de nuestros amigos. Consideramos que si aceptamos que la intención de voto es un acto con de naturaleza contagiosa y las redes sociales en Internet son capaces de manifestar características propias –la influencia de nodos no centrales, la longitud y forma de la cadena de difusión, multiplicación de conexiones, proliferación de vínculos débiles- que difieren de las teorizadas en las redes tradicionales, en este canal, el contagio político puede adquirir valores y comportamientos que sorprenderían por su cantidad y procedimientos únicos. Es por esto que el análisis de las actitudes sociales y fenómenos de difusión y contagio en estas redes, constituyen un área del mayor interés para sociólogos, politólogos, asesores, estrategas e investigadores, entre otros, por su enorme implicación tanto en estudios socio-demográficos como en estudios de comportamiento electoral y definiciones de estrategias en campañas electorales.

Además de revisar algunas teorías sobre el mecanismo de contagio político entre individuos, nos preguntamos por la capacidad de éxito en el uso de redes sociales en una campaña electoral y cuáles son las líneas de actuación que marcan la diferencia. ¿Cómo decidió la gente sumarse a la campaña de Barack Obama en Estados Unidos en 2008 cuando todavía no se había extendido la percepción del candidato como posibilidad? ¿Cómo logró persuadir a tanta gente que hasta entonces no se había implicado en política? Obama triunfó en gran medida porque muchas personas se sintieron conectadas. Se ha comentado la extraordinaria capacidad de Obama de conectar con los votantes, pero más impresionante aún fue su capacidad de conectar a unos votantes con otros.

El análisis de campañas electorales online ha manifestado que la red tiene capacidad de influir en un efecto positivo en los resultados electorales y que los candidatos con mejor posicionamiento en las redes sociales pueden aventajar a sus oponentes. Pero la comprensión de la realidad no es tan básica ni fácil de prever en este campo.

En lo que respecta a la polarización de grupos u homofilia extrema que se venía atribuyendo como característica de comportamiento en Internet corroborada en el estudio sobre la blogosfera americana en las elecciones de 2004, se ha constatado que al menos en la red Twitter, la subred de “retweets” aparece como altamente partidista mientras que la subred de “menciones” se muestra como un grupo homogéneo en el que individuos con divergencias ideológicas conversan entre sí.

Con la sucesión de estudios dedicados al análisis de estas Redes Sociales, hoy sabemos más acerca de la forma en que las difusiones y los contagios se suceden en ellas. No obstante, y dada su complejidad y novedad, se hace patente la necesidad de profundizar en este campo, con investigaciones continuadas en el tiempo y diversas en sus objetivos. Futuros análisis comparados, tanto de naturaleza cualitativa como cuantitativa ayudarán a conseguir conclusiones certeras más cercanas a la realidad. Es aquí dónde esta investigación pretende con sus métodos y resultados, aportar un ápice de conocimiento a este nuevo y relevante campo de estudio.

Capítulo 2

Marco teórico

3. Capítulo II: Marco Teórico

3.1. Introducción

En nuestro marco teórico se han estudiado los temas que se han considerado relacionados con distintos aspectos que atañen a las conclusiones de esta investigación. En su conjunto, se trata de una serie de datos estadísticos y estudios de investigación con objeto similar al que nos hemos propuesto. Alguno de estos trabajos han sido el origen y la inspiración de esta Tesis Doctoral, y otros han servido para extraer piezas claves de conocimiento probado que se incorporarán a la metodología empleada.

Los temas en los que nos hemos fijado son, por orden de aparición en el capítulo, los siguientes:

- “El auge de la política en Internet”, epígrafe en el que se pone de manifiesto tanto el incremento del uso de Internet por parte de los usuarios para informarse sobre diversas opciones políticas como la intensificación de la presencia de candidatos y partidos en la Red, sobre todo para divulgación y promoción de sus campañas electorales.
- “Antecedentes del uso del análisis de Redes Sociales como herramienta de predicción”, apartado dedicado a la exposición de estudios realizados mediante el análisis de la actividad de los usuarios en Internet con el objeto de predecir las tendencias de opinión pública.
- Tres epígrafes -referidos a “Facebook, Google y Twitter como herramientas de predicción electoral”-, en los que recorren los estudios publicados por diferentes autores que han vigilado distintas contiendas electorales a través de la monitorización de los datos que los usuarios de estas redes vertían en cada una de ellas.
- “El peso de las Ciber campañas en el éxito electoral”, en el que se evidencia el mayor o menor peso que la dotación de una Ciber campaña

puede aportar de cara a la captación de votos y por tanto, a la mejora del resultado electoral.

- “Encuestas preelectorales. Eficacia y desviación”, dedicado a constatar la realidad de las encuestas preelectorales como métodos fiables de predicción de los resultados electorales.
- “Grado de penetración de las Redes Sociales en la población española”, destinado a la obtención de datos característicos de los usuarios de las Redes Sociales en España.

Los dos primeros, “El auge de la política en Internet” y “Antecedentes del uso del análisis de Redes Sociales como herramienta de predicción”, cumplen una función de carácter contextualizador y de puesta en escena de la situación en la que queda enmarcada nuestra investigación. Al mismo tiempo, se indica la conexión entre la política en red y el análisis de Redes Sociales en Internet para esta especialidad.

Los apartados referidos particularmente a “Facebook, Google y Twitter como herramientas de predicción electoral” junto con el titulado “El peso de las Ciber campañas en el éxito electoral”, además de referenciar y contrastar los estudios publicados por varios autores en este campo, han servido como fuente de variables y consideraciones que serán aplicadas y/o tenidas en cuenta en el desarrollo de la metodología de análisis.

“Encuestas preelectorales. Eficacia y desviación” sirve a nuestra investigación ya que no sólo compararemos nuestros resultados con los resultados electorales definitivos, sino que cotejaremos los datos de sus desviaciones con las que se midan en este estudio.

Por último, en el “Grado de penetración de las Redes Sociales en la población española” se extraen datos dirigidos a determinar en nuestra investigación si los usuarios de estas redes pueden considerarse como una muestra representativa de un colectivo poblacional determinado en una elección dada.

3.2. El auge de la política en Internet

El consumo de “política” en la red es un hecho ya probado. Las personas consumen o empiezan a consumir noticias o a seguir con un cierto entusiasmo las elecciones vía Internet. En el estudio “The Internet and the 2008 Election” realizado por *Pew Internet & American Life Project* tras la celebración de las elecciones en Estados Unidos, se manifestó que cerca de un 46% de ciudadanos ha empleado Internet para informarse sobre la campaña electoral, compartir sus opiniones y movilizar a otros (Smith y Rainie, 2008, p.3). Y la tendencia no se resume a la campaña estadounidense. En el Informe sobre la blogosfera hispana que cada año edita Bitacoras.com⁹⁷ que indexa cerca de 500.000 blogs, en el año 2011 los dedicados al tema político ocupaba el puesto número 6 en un rango de 22 tipologías de vertientes editoras. Según una encuesta publicada en la página Web del parlamento Europeo los ciudadanos consideran beneficioso el uso de las tecnologías en las campañas electorales, concretamente un 57% opina favorablemente acerca de su efecto político⁹⁸. En España, la Encuesta sobre Equipamiento y Uso de Tecnologías de la Información y Comunicación en los hogares 2011 (INE 2011) revela que un 21,2% de los Internautas ha accedido a Internet para leer y emitir opiniones sobre asuntos de tipo social o político en lugares de la Red.

En el primer uso de Internet como herramienta de campaña desplegado por el candidato demócrata a la presidencia de Estados Unidos en 2004, Howard Dean, el trabajo social en red aún no se había inventado. Pasarían cuatro años hasta que la candidatura de Barack Obama utilizará Internet y las redes sociales para pasar de ser el candidato menos probable, sin mucho dinero y con pocos apoyos, a vencedor en las primarias frente a Hillary Clinton y presidente de Estados Unidos derrotando en las urnas al candidato republicano John McCain. Su estrategia en la red marcó un

⁹⁷“Informe sobre el estado de la blogosfera hispana”. 2011. *Bitacoras.com* <http://bitacoras.com/informe/11#tematicas>. (Consultado el 12 de marzo de 2012)

⁹⁸ “Las redes sociales, herramientas útiles para las elecciones europeas” Mayo de 2009. *elmundo.es* <http://www.elmundo.es/elmundo/2009/05/21/navegante/1242900454.html> (Consultado el 10 de marzo de 2012)

antes y un después en la forma de las campañas electorales y se asegura que su elemento diferenciador es que en ella “triunfó la comunidad”.

Si bien es cierto que Obama destinó una cantidad mayor de recursos en Internet que sus oponentes, también lo es que gastó 10 veces más, según los datos oficiales de la *Federal Election Commission*, que Hillary Clinton en publicidad digital en Google, Yahoo y Facebook. Además, otros factores contribuyeron a su victoria: la capacidad de organizar a ciberactivistas que trasladan su actividad al “mundo real” y el hecho de que el 74% de los seguidores online de Obama han consultado información política online comparado con el 57% de los seguidores online de Clinton y también han superado a los seguidores online de McCain (65% vs. 56%), habiendo éstos efectuado más contribuciones campaña, más firmas para peticiones online, más comentarios en blogs y más vídeos vistos que el resto del electorado. (Smith and Ranie, 2009)

¿Cómo decidió la gente sumarse a ella cuando todavía no se había extendido la percepción de Obama como posibilidad? ¿Cómo logró persuadir a tanta gente que hasta entonces no se había implicado en política? *Obama triunfó en gran medida porque estos “hombres y mujeres trabajadores” se sintieron conectados. (...) Muchos han comentado la extraordinaria capacidad de Obama de conectar con los votantes, pero más impresionante aún fue su capacidad de conectar a unos votantes con otros.* (Christakis y Fowler, 2010, p.185)

Los seguidores de Obama se mostraron más activos en la red que los seguidores de sus dos oponentes. Un 74% de los seguidores online de Obama han consultado información política online comparado con el 57% de los seguidores online de Clinton, y comparados con los seguidores online de McCain, los seguidores de Obama también ganaron con un 65% vs. un 56%. También sus seguidores superaron al resto en participación en acciones directamente relacionadas con la campaña online, concretamente comparados con los seguidores de Clinton, hicieron más contribuciones campaña (17% vs. 8%), firmaron más peticiones online (24% vs. 11%), dejaron más comentarios en blogs y otras páginas Web (23%

vs.13%), y consumieron más vídeos en la red (64% vs. 43%)⁹⁹. (Smith y Rainie, 2008, p.6)

A partir de esta campaña, la mayoría de los políticos que se debaten en un proceso electoral utilizan Internet y sus redes como canal de marketing electoral para difundir sus propuestas, contactar y dialogar con los ciudadanos, y en suma, para ganar las elecciones y conseguir el poder (y mantenerlo). César Calderón destaca algunos casos significativos (Calderón et al., 2010, pp.172-176)

- El presidente de Ecuador, Rafael Correa, fue el primer presidente latinoamericano que abrió su propio blog (El Ciudadano Presidente Rafael Correa).
- La plataforma “Súmate a Cristina”, primera cibercampaña presidencial lanzada en Argentina, ideada para difundir los mensajes de la candidata y actual presidenta Cristina Fernández de Kirchner en las elecciones de 2007.
- La contienda Sebastián Piñera-Eduardo Frei en las elecciones 2010 en Chile y sus diferencias online. Piñera contó con varias páginas Web's dirigidas a diferentes segmentos, su página en Facebook alcanzó seis veces más seguidores que la de Frei y fue el único que aprovechó Twitter. Piñera ganó las elecciones.

Ejemplos como estos nos sugieren que en campaña electoral, estar en la red tiene capacidad de influir en un resultado positivo en los resultados electorales y que los candidatos con mejor posicionamiento en las redes sociales pueden aventajar a

⁹⁹*El equipo de video de Obama está liderado por Kate Albright-Hanna que ha sabido explotar el potencial de YouTube para difundir el mensaje audiovisual de Obama. Los más de 1300 videos (y más de 2000 horas, a mediados de Septiembre de 2008) superan ampliamente a los disponibles (menos de 300) en el canal de John McCain. Pero, lo que es más relevante es que el impacto de los videos de Obama ha sido muy superior (unos 88,000 suscriptores y 16 millones de visitas al sitio de YouTube respecto a unos 20,000 suscriptores y poco más de 1,4 millones de visionados).* (FREIRE, Juan. 09/11/2008. Influencia de Internet en las elecciones en EEUU. Internet. <http://nomada.blogs.com/jfreire/2008/11/influencia-de-internet-en-las-elecciones-en-eeuu-cuadernos-de-periodistas.html>. (Consultado el 03/03/2009)

sus oponentes. Pero la realidad no es tan básica ni fácil de prever en este campo. Antanas Mockus, candidato electo a la presidencia de Colombia por el Partido Verde era el preferido en redes sociales como Facebook¹⁰⁰ o Twitter¹⁰¹ y sin embargo fue derrotado por su oponente Juan Manuel Santos, consiguiendo, no obstante, ser la segunda fuerza política del país con más de tres millones de votos recabados por su causa en las urnas y siendo éste su mejor resultado electoral.

Javier Duarte de Ochoa, candidato en 2010 por el partido PRI a la gobernatura del Estado de Veracruz en México alcanzaba los mayores índices de popularidad en la red a pesar de que *No tiene el mayor número de seguidores en Twitter ni es el de mayor nivel de influencia global en esa red social, ese honor corresponde a Xóchitl Gálvez (Bloque opositor, Hidalgo); no es el que acumuló comparativamente nuevos seguidores en Twitter en la semana, ése es Antonio Mejía (PRD, Zacatecas); no es el rey de los amigos/fans en Facebook, ése es Mario López (Bloque opositor, Sinaloa); tampoco es el que reporta la mayor cantidad de resultados en Google, ese honor corresponde a su rival Miguel Ángel Yunes*¹⁰². Ello se debió a que en todas las categorías analizadas por los autores del seguimiento de la campaña en Internet, resultaba claro vencedor. El candidato resultó vencedor en las urnas por un reducido número de votos.

Asumimos que el resultado de cualquier disputa electoral no puede fijarse, únicamente, en la calidad de la presencia del candidato en las redes sociales en Internet, ya que entran en juego muchos otros factores como la imagen, credibilidad y liderazgo del aspirante, el clima social y la oportunidad de conseguirlo, o el éxito

¹⁰⁰ La página de Mockus en Facebook, <http://www.facebook.com/a.mockus>, cuenta con el doble de seguidores que la del vencedor Santos,, <http://www.facebook.com/a.mockus#!/JMSantos.Presidente?v=wall&ref=ts>, (791,719 vs.369,128) (Consultado el 28/07/2010)

¹⁰¹ Mockus cuenta con 65,855 seguidores. No se ha encontrado referencia sobre la cuenta en Twitter de Santos. (Consultado en Internet el 28/07/2010)

¹⁰² José Merino, Eduardo Fierro y Raúl Valencia.UIEE.13/10/2010. "Candidatos opositores, los más atacados en línea". Internet. <http://eleconomista.com.mx/elecciones-2010/2010/06/13/candidatos-opositores-mas-atacados-linea>. (Consultado el 15/06/2010)

de las campañas físicas que se desarrollan a la par, pero intuimos que influye en los resultados. Además, hemos visto cómo candidatos con mejores índices de popularidad en la Red no han visto reflejada tal ventaja en el resultado electoral obtenido, como el caso de Antanas Mokcus, anteriormente citado. También podemos citar ejemplos en los que candidatos que no han desarrollado actividad web alguna, han resultado victoriosos en la contienda electoral, como es el caso del cercano Ramón Luis Valcárcel en las pasadas Elecciones Autonómicas de la Región de Murcia, caso en el que se ha de tener en cuenta, entre otras variables, su posición de reelección, la cobertura recibida por los medios de comunicación tradicional, y las características ideológicas y demográficas del electorado.

En los últimos años, los medios y redes sociales se han convertido en un espacio del mayor interés para la colaboración social y compartir información entre sus usuarios. La ingente cantidad de contenido que circula en ellas ha atraído la atención de agencias de marketing que esperan conocer el comportamiento de los usuarios para ajustar sus campañas online y de investigadores en el campo de las redes sociales que buscan en sus análisis revelar los modelos de difusión entre grupos y nodos en dichas redes, e incluso demostrar como ese contenido social puede ser utilizado para predecir resultados en el mundo real (Asur y Huberman, 2010). Recientemente este tipo de análisis ha saltado al campo de la política, y especialmente, al análisis de campañas electorales en la red y concretamente, al uso de métodos que intentan predecir el resultado electoral monitorizando la interacción que candidatos y electores mantienen en este canal.

En el año 2010, un estudio del equipo de investigación de la consultora tecnológica *Gartner*, reveló las diez tecnologías estratégicas para el 2011¹⁰³. Entre ellas se citó el Análisis de Redes Sociales, que se describió como el proceso de

¹⁰³ SELLERS, Dennis. (2010) "Apple and the top 10 'Strategic Technologies' for 2011" <http://www.macnews.com/2010/10/27/apple-and-top-10-strategic-technologies-2011/>. (Consultado el 08/11/2010)
<http://www.gartner.com>

medir, analizar e interpretar los resultados de las interacciones y asociaciones entre las personas, temas e ideas. Incluye una serie de técnicas de análisis especializadas como el filtrado social, el análisis de redes sociales, el análisis de los sentimientos que circulan por ellas y el análisis de los medios sociales empleados. Las herramientas de análisis de redes sociales son útiles para examinar la estructura social y la interdependencia, así como las pautas de comportamiento de los individuos, grupos u organizaciones. El análisis de redes sociales incluye la recolección de datos de múltiples fuentes, identificando las relaciones, y evaluando el impacto, calidad o eficacia de una relación.

La idea de que lo que las personas están publicando en Internet o buscando en ella puede aportar una idea del sentimiento colectivo junto con el hecho de que la mayoría de los medios sociales en Internet ponen a libre disposición de cualquiera el uso de unas API's con las que recolectar los datos publicados por sus usuarios, ha provocado que investigadores y periodistas vean en el análisis de redes sociales la posibilidad de tomar el pulso de la opinión pública.

En el campo de la ciencia política y concretamente en lo que se refiere a la capacidad de predecir un resultado electoral mediante el estudio de redes sociales, destaca el análisis de datos en Twitter tanto por ser una red extendida y reconocida como medio de expresión y difusión de opiniones, como por la facilidad que ofrece para capturar dichos datos gracias a su API de libre disposición y explotación que permite recolectar lo que se conoce como *public time-line*, esto es todos los tweets de sus usuarios salvo los de aquellos que mantienen un perfil privado que protege sus tweets del resto. Además de Twitter, también podemos encontrar estudios que relacionan el número de seguidores o fans de una página Facebook de un candidato con su resultado electoral, y otros que han utilizado el volumen de búsquedas realizadas en Google sobre una opción política determinada.

Teniendo en cuentas los principales medios sociales en Red en los que hoy en día, al menos en España, se articulan las Ciber campañas Electorales, podemos contemplar varios espacios posibles para el análisis del impacto de una campaña en

la Red: en Facebook, Twitter, Google Insights y la página Web de partido y/o de candidato. Gracias a la forma de la propia Red, es posible conseguir y utilizar los rastros de datos que estas campañas dejan en la misma, a modo de datos cuantitativos capaces de servir como elementos de predicción electoral, aplicando distintas metodologías de análisis según la Red Social (Twitter por ejemplo) o elemento Web empleado en el análisis (Google Insights for Search) de que se trate.

En los siguientes apartados mostraremos los avances conseguidos en este campo para cada una de las herramientas Web mencionadas, basándonos en estudios publicados por diversos autores especialistas en ciencia política y en análisis de redes.

3.3. Antecedentes en el uso del análisis de Redes Sociales como herramienta de predicción

En estos años recientes, los medios sociales en la Red se han desvelado como un espacio relevante para la colaboración social y la difusión compartida de contenidos, opiniones y valoraciones objetivas y particulares sobre cualquier tema que conviva con nosotros en esta sociedad, y que nos afecta y/o interesa en uno u otro sentido. Desde opiniones sobre el teléfono móvil que mejores prestaciones ofrece, pasando por el hotel que reúne la mejor relación calidad/precio hasta nuestras preferencias políticas y sus razones. Internet y sus herramientas de colaboración social ponen a nuestra disposición todo un universo de valoraciones compartido de forma universal, sin más límite (a priori) que la diferencia de códigos impuesta por las diferencias lingüísticas.

El estudio sobre la influencia en los *social media* de la consultora McCann, bajo el curioso título “When did we start trusting strangers?”, revela que ahora todos podemos influir en los demás, que ya no sólo los medios de comunicación o la clase social dominada por especialistas en un área concreta, pueden influir en nuestra toma de decisiones. Afirman, además, haber encontrado una nueva clase de influyentes a

los que denomina “super-influyentes”. Individuos que gracias a los social media y a las herramientas que proveen para crear y compartir juicios, superan el grado medio de la influencia difundiendo su opinión más lejos y más ampliamente. Son personas inteligentes, con un alto nivel de educación y un profundo conocimiento de Internet y ordenadores. Estas personas generan y comparten contenido, y se ajustan al perfil de *early-adopters* (adaptadores tempranos) dispuestos a probar nuevos productos, asumir riesgos y compartir sus opiniones con amigos y con extraños. La consultora afirma que la actividad de esta clase de “super-influyentes” fue determinante en la victoria electoral de Barack Obama en las elecciones de 2008 en Estados Unidos: *It has been widely credited that the creators or creative class, as they have been tagged, helped Obama get the democratic party nomination. The creators through prolific blogging, twittering, social networking and content creation generated massive influence helping to raise funds, getting people out to vote and changing the opinions of mainstream media.* (Universal McCann, 2008, pp.55-56)

Desde la aparición de los medios o redes sociales en Internet, sus desarrollos creativos en torno a estructuras que faciliten la conexión y difusión común en Red no han dejado de crecer: herramientas sofisticadas que permiten a sus usuarios compartir sus opiniones y/o sentimientos sobre un tema particular, etiquetados que aportan un gran valor añadido a las informaciones que publican al convertirlas en objetivos fáciles de encontrar por similitud de gustos o tendencias, y nuevas estrategias de conexión entre perfiles en la Red que permiten conectar a millones de personas por afinidades tales como demográficas, sociales, culturales e ideológicas.

Dos investigadores de HP Labs, en Palo Alto, California, pensaron que dada las ingente cantidad y variedad de información que transita a través de extensas comunidades establecidas en torno a las redes sociales, sería factible tomar como fuente de análisis todo este caudal de información que circulaba gratuitamente y de forma accesible por Internet, con sus riquezas de etiquetado extensivo y variado sobre la información que por ella se mueve, y que por tanto las conversaciones multitudinarias podrían ser analizadas y utilizadas para predecir las preferencias del

mercado, en un momento dado. En suma, valerse del análisis de redes sociales para predecir una tendencia económica-social-política cualquiera: *“Moreover, gathering information on how people converse regarding particular products can be helpful when designing marketing and advertising campaigns.”* (Asur y Huberman, 2010, p.1)

El objeto de su estudio fue concluir si la monitorización de la conversación sobre determinadas películas o films en Twitter sería capaz de predecir el éxito en taquilla de todas o cualquiera de ellas, vigilando el volumen de datos en Twitter y su propagación (*Twitter mentions*), y desarrollando un método de análisis sentimental (*Sentiment Analysis*) al efecto. Tras el estudio, que recolectó 2.89 millones de tweets referidos a 24 películas, se trataba de comparar sus resultados con los datos producidos por una “agencia de valores” (películas y actores) en Internet, *Hollywood Stock Exchange*¹⁰⁴, concluyendo que su análisis era capaz de predecir resultados de venta de taquilla mejor que los métodos tradicionales: *“Moreover our predictions are consistently better than those produced by an information market such as the Hollywood Stock Exchange, the gold standard in the industry.”* (Asur y Huberman, 2010, p.1)

Sus conclusiones parciales y de gran interés para el análisis de este tipo de datos, son las siguientes.” (Asur y Huberman, 2010, pp.3-7):

- Atención al lanzamiento. El momento en que las compañías y productores lanzan los trailers (vídeos, blogs y fotos) de una película, se produce una correlación de menciones y RT’s en Twitter. Cuanta más publicidad mayor repercusión en la recaudación en taquilla.
- La antesala. Analizando los tweets previos al lanzamiento comercial, es posible predecir al futuro comercial de la misma¹⁰⁵.

¹⁰⁴ *Hollywood Stock Exchange* “is The Entertainment Market™ where you can buy and trade movies, stars and more”. <http://www.hsx.com/>

¹⁰⁵ Utilizaron la formula “tweets-rate” igual al número de tweets publicados por hora referidos a una película.

- Más allá del lanzamiento. Analizando los tweets en cualquiera de los fines de semana monitorizados, sus resultados concluyeron que midiendo cualquiera de estos fines de semana su método era mejor predictor de las recaudaciones de ese fin de semana que cualquiera de los métodos tradicionales existentes.
- En el “sentiment analysis” obtuvieron una precisión del 98%. Además, encontraron que el número de sentimientos positivos hacia las películas es mayor que el negativo y que este análisis no produce mejoras en comparación con el análisis único de los tweets.
- La subjetividad aumenta una vez terminado el periodo de mantenimiento de película en cartel.

También con datos recogidos en la red Twitter, Zhang, Fuehres y Gloor (2010) se interesaron por la capacidad predictiva del análisis de los datos de esta red. En su artículo tratan de predecir los indicadores bursátiles como el Dow Jones, Nasdaq y S&P 500 mediante el análisis emocional de los mensajes de Twitter. Se recogieron los *feeds* de Twitter durante seis meses y se eligió una submuestra al azar de aproximadamente una centésima parte del volumen total de todos los tweets. Se midió la esperanza colectiva y el miedo cada día¹⁰⁶ y se analizó la correlación entre estos índices y los indicadores del mercado de valores. Se encontró que el porcentaje de tweets emocionales estaba negativamente correlacionado con Dow Jones, Nasdaq y S&P 500, pero mostraba una correlación positiva significativa con respecto al índice VIX¹⁰⁷. Concluyen que la sola comprobación en Twitter de las muestras emocionales de todo tipo pueden convertirse en un indicador de cómo el mercado de valores va a comportarse al día siguiente: *“To put it in simple words, when the emotions on twitter fly high, that is when people express a lot of hope, fear, and worry, the Dow goes down the next day. When people have less hope, fear, and worry, the Dow goes up. It therefore seems that just checking on twitter for*

¹⁰⁶ Tweets que contienen las palabras: *hope, fair* y *worry*.

¹⁰⁷ The Chicago Board Options Exchange Volatility Index VIX.

emotional outbursts of any kind gives a predictor of how the stock market will be doing the next day.” (Zhang et al., 2010, p.7)

Aproximaciones similares fueron elaboradas por investigadores de la Universidad de Indiana, logrando un método para predecir los cambios en el *Dow Jones Industrial Average* (DJIA) a través del análisis de los estados de ánimo recogidos en las actualizaciones de los usuarios de Twitter. Bollen, Mao y Zeng (2010) basaron su investigación en el hecho de que el comportamiento en lo que respecta a la economía, nos dice que las emociones pueden afectar profundamente el comportamiento individual y la toma de decisiones. ¿Puede aplicarse de igual forma a las sociedades en general, es decir, las sociedades pueden experimentar cambios en sus estados de ánimo que influyan en su toma de decisiones colectivas?

Con el uso de dos algoritmos de registro del estado de ánimo, el *Google-Profile of Mood States* (GPOMS) y *OpinionFinder*, los investigadores analizaron 9,7 millones de *tweets* publicados entre marzo y diciembre de 2008, encontrando que las correlaciones para el índice de tranquilidad (*calmness*), uno de los seis estados de ánimo medido por GPOMS, podría ser utilizado para predecir si el DJIA aumentó o disminuyó entre dos y seis días más tarde. Sus resultados indican que la exactitud de las predicciones del DJIA puede mejorarse significativamente por la inclusión de las dimensiones específicas de un estado público de ánimo (*calmness*) pero no en otros, encontrando una precisión del 87,6% en la predicción diaria en los valores de cierre del *Dow Jones* y una reducción del MAE (*Mean Average Percentage Error*) en más de un 6%.

Lo interesante de este seguimiento de los estados de ánimo de Twitter es que de los seis estados GPOMS medidos - felicidad, bondad, estado de alerta, seguridad, vitalidad y tranquilidad – la tranquilidad es el estado de ánimo más útil en la predicción de los cambios del mercado de valores. Ninguno de los otros índices, incluyendo los indicadores de sentimiento positivo / negativo que facilita la herramienta *OpinionFinder*, refleja todos los cambios del mercado de valores: “*Surprisingly we do not observe this effect for OpinionFinder’s assessment of public*

mood states in terms of positive vs. negative mood but rather for the GPOMS dimension labeled “Calm”. The calmness of the public (measured by GPOMS) is thus predictive of the DJIA rather than general levels of positive sentiment as measured by OpinionFinder.” (Bollen et al., 2010, p.6)

Dado que los métodos de vigilancia tradicionales (encuestas) que existen en el mercado para averiguar o predecir las tendencias de futuro de la población procuran los resultados de sus investigaciones con un cierto retraso sobre el momento en que están realmente sucediendo los acontecimientos que se persiguen conocer, la inmediatez y el bajo costo de la herramienta de Google también ha empezado a ser foco de atracción para investigadores de diversos ámbitos del consumo (industria del automóvil, preferencias de turismo, ventas de casas), del control de la salud pública (seguimiento de epidemias) y, recientemente, de tendencias de voto electoral (predicción electoral en campaña)¹⁰⁸.

En “Predicting the Present with Google Trends” (Choi y Varian, 2009) se evalúa si Google Trends puede ser un predictor de la actividad económica. Los autores compararon los resultados de búsqueda entre los años 2004 y 2008 acerca de varios sectores de la economía de los Estados Unidos de América con los informes publicados por US Census Bureau, US Department of Housing and Urban Development y Hong Kong Tourism Board, en los ítems Ventas de automóviles, Ventas de casa y Viajes. Sus resultados, que arrojaron MAE’s de diferentes porcentajes (variando entre un 5% y un 11%), les llevaron a la conclusión de que si bien en ese momento Google Trends no podía ser tomado como un confiable predictor de la actividad económica, sí podía convertirse en una herramienta de ayuda: *“We are not claiming that Google Trends data help predict the future. Rather we are claiming that Google Trends may help in predicting the present. For example, the volume of queries on a particular brand of automobile during the second week in June may be helpful in predicting the June sales report for that brand, when it is*

¹⁰⁸ Dado el objetivo de este trabajo, las investigaciones dedicadas a predicción electoral con Google se incluyen en el apartado específico dedicado a este campo.

released in July”. (Choi y Varian, 2009, p.II). Los autores advierten de que la forma en la que Google Trends computa la toma de muestras varía día a día, por lo que estas variaciones podrían haber introducido “ruido” en sus datos, por lo que son conscientes de que su modelo puede servir de base para futuros estudios más sofisticados que podrían mejorar el uso apropiado de esta herramienta como indicador de tendencias sociales.

Las epidemias de influenza o gripe estacional son una de las grandes preocupaciones en salud pública. Ginsberg et al. (2009) pensaron que una forma de mejorar la detección temprana sería monitorear la salud en base a la conducta de búsqueda que millones de personas en todo el mundo realizan en Internet sobre este tema cada día. Los autores basan su hipótesis en que debido a que la frecuencia relativa de determinadas consultas en la Red está altamente correlacionado con el porcentaje de visitas al médico en el que un paciente se presenta con síntomas gripales, se podría estimar con precisión el nivel actual de la actividad semanal de la gripe, con un desfase de información de aproximadamente un día, un rango inferior al que presenta el CDC (U.S Center for Disease Control and Prevention) que alcanza una posterioridad de dos semanas: *“About 90 million American adults are believed to search online for information about specific diseases or medical problems each year⁷, making web search queries a uniquely valuable source of information about health trends¹⁰⁹.”* (Ginsberg et al., 2009, p.2)

En su método agregaron las búsquedas realizadas durante los años 2003-2007 contabilizando series temporales semanales de más de 50 millones de las búsquedas más comunes en los EEUU. Cada búsqueda fue normalizada dividiendo el número de búsquedas por una *query* particular entre el total de búsquedas realizadas esa semana en una región concreta. Los *topics* o términos de búsqueda utilizados en su investigación fueron aquellos encontrados automáticamente en Google que

¹⁰⁹POLGREEN, Philip M., CHEN, Yiling, PENNOCK, David M. y NELSON, Forrest D. (2008) “Using internet searches for influenza surveillance”. *Clinical Infectious Diseases*, 47, 1443–1448. Mencionado por Ginsberg et al. (2009)

mantenían una relación directa con los datos publicados en ILI¹¹⁰. Los resultados de su investigación fueron realmente positivos en esta ocasión¹¹¹: *“Across the nine regions, we were able to consistently estimate the current ILI percentage 1-2 weeks ahead of the publication of reports by the CDC’s U.S. Influenza Sentinel Provider Surveillance Network.”* (Ginsberg et al., 2009, p.3)

También en esta ocasión y coincidiendo con las recomendaciones que veremos en los estudios recogidos en siguientes apartados de este marco teórico dedicados a Facebook y a Twitter, los autores nos advierten, además de que su método puede estar sometido a variables no controladas como falsas alarmas que desatan una euforia en ciertas búsquedas, de que estos métodos no pueden ser utilizados en sustitución de los métodos tradicionales de vigilancia de la salud, entre otras cuestiones, por la inexistencia de la riqueza que ofrece la obtención y el estudio de datos demográficos detallados: *“This system is not designed to be a replacement for traditional surveillance networks or supplant the need for laboratory-based diagnoses and surveillance.(...) Demographic data, often provided by traditional surveillance, cannot be obtained using search queries.”* (Ginsberg et al., 2009, p.3)

3.3.1. Facebook como herramienta de predicción electoral

Las características de privacidad de la red social Facebook dificultan la labor de análisis. Si bien en el análisis de datos en Twitter, contar con variables externas a la propia red, referidas a las características intrínsecas de la campaña y de sus candidatos, incrementaría la veracidad de los resultados de la investigación, en Facebook, esta realidad es prácticamente una obligación de entrada. Los análisis de una campaña electoral en esta red que hasta ahora se han intentando y difundido, articulan su criterio sobre el número de “like” equivalente al número de “supporters” o “seguidores” que haya obtenido cada opción política en su página Facebook

¹¹⁰ *Regional and state-level estimates of influenza-like illness.*

¹¹¹ Este método es empleado para rastrear el estado de la gripe en todo el mundo. Puede verse en <http://www.google.org/flutrends/>. (Consultado el 28/04/2012)

durante el tiempo de campaña electoral. Esto es, en su mayor parte, debido a su carácter de privacidad, por el que los usuarios de esta red social sólo comparten sus datos (mensajes, comentarios, likes, etc.) con aquellos contactos de su red con quienes desean hacerlo, de forma que no es posible monitorizar las conversaciones y/o menciones que se producen sobre un candidato a lo largo y ancho de la misma. Así, para el análisis de campañas electorales, el único dato con que se cuenta para la captura de datos es la página de perfil público o “página fan” de candidatos y partidos, de la que puede extraerse información referida a la evolución de seguidores, comentarios publicados por el propio perfil, *post* de sus seguidores, etc.

Con motivo de las elecciones de 2006 en EEUU, Facebook creó un espacio especial en su red, *US Politics*, en el que podrían almacenarse los perfiles de los candidatos, diferenciándose así de las páginas personales. Estas nuevas páginas de perfiles políticos partían de una plantilla que contenía el nombre del candidato e información básica de contacto. Las *passwords* que permitían a los candidatos asumir la responsabilidad de personalizar y actualizar su perfil eran enviadas por Facebook a las oficinas de los Partidos Demócrata y Republicano, desde dónde se distribuían a los candidatos finales. Facebook se convertía así en la red social líder en campañas electorales. Dos años más tarde, en el contexto de las elecciones de 2008, Facebook introdujo nuevas modificaciones para las páginas de políticos y/o candidatos, permitiendo la publicación en ellas de nuevos tipos de contenidos (enlaces, vídeos, álbumes de fotos, etc.) y extendiendo su uso al resto del mundo. Se eliminó el espacio “US Politics” y todas las páginas de los candidatos fueron ubicadas bajo la sub-sección “Politician”, junto a las Fan Pages de celebridades y figuras públicas, como artistas, deportistas y productos, entre otras categorías. A pesar de de estas mejoras y del aumento de la visibilidad de las Fan Page de políticos, sólo el 8% de los americanos mencionó haber usado las redes sociales para informarse acerca de la campaña de 2008 (Smith y Ranie, 2008).

Facebook empezó en ese momento a convertirse en un espacio Web Social en el que políticos de todo el mundo participaban con la creación y publicación de su

propia página Fan. Era estar y darse a conocer o no estar y perder cierta probabilidad de victoria.

“Nationwide this year social networking has been huge, and that is becoming obvious. Anyone in politics who tries to ignore that is foolish quite frankly, and is going to have no chance at continuing to win elections. «Ryan», campaign staff of Roger Wicker (R-MS).” (Willians y Gulati, 2009, p.3)

Y su espacio político continuó creciendo. Una de las razones por las que redes sociales como Facebook (o Twitter) son ahora tan populares en todo el mundo es el hecho de que, con el desarrollo de la conectividad móvil, ya no es necesario estar “enchufado” en el salón de tu casa para poder participar en ellas: *“Most Americans who became fans of Obama or Brown probably did so via a desktop computer; in places like the Philippines, though, more almost certainly fanned their candidates through mobile devices.” (U.S Politics on Facebook, 2010)*

En noviembre de 2010 la división política de Facebook, *U.S. Politics on Facebook*, emitía una nota en la que aseguraba que según sus datos en esta red la mayoría de candidatos con más seguidores habían vencido a sus oponentes en las elecciones legislativas de 2010: *“The Facebook political team’s initial snapshot of 98 House races shows that 74% of candidates with the most Facebook fans won their contests. In the Senate, our initial snapshot of 19 races shows that 81% of candidates with the most Facebook fans won their contests.” (U.S Politics on Facebook, 2010)*

En las mismas fechas, año 2010, la empresa de consultoría Web Trilogy que ha asesorado en las redes sociales a decenas de candidatos en campaña durante más de cuatro años, advierte que existe una estrecha relación entre Facebook y la victoria electoral, aunque dicha correlación no se da en todas las campañas estudiadas. En su trabajo “Facebook predicts electoral victory or not?” los autores afirman que: *“We looked at the data to see if the total number of Facebook Likes for a given candidate was a good predictor of victory on Election Day. Our results show some correlation*

in U.S. Senate races, but that correlation effectively disappeared in U.S. House and gubernatorial races” (Olson y Bunnett, 2010, p.2)

Aunque los medios de comunicación social por sí solos no ganan las elecciones, las plataformas como Facebook se han convertido cada vez más una espacio integral para emitir opiniones sobre los candidatos de todo el mundo, y mientras que el número de fans de Facebook no puede traducirse directamente en número de votos, parece encontrarse al menos algún tipo de correlación entre ambos.

Por otra parte, si bien Cha et al. (2010) concluyeron en su estudio sobre Twitter que el número de *followers* en esta red no es indicador de la capacidad de influencia de un individuo sobre los componentes de su red y que por tanto los mensajes que ellos emitían no calarían ni serían re-difundidos o “retuiteados” necesariamente por sus seguidores alcanzando así nuevas audiencias con capacidad de influir a su vez en sus propios seguidores y quizás contagiar a terceros con las opiniones vertidas en primera línea de emisión, en el caso de candidatos y partidos políticos, bien podríamos aventurar que aunque sus mensajes (ideologías y/o fragmentos claves de sus programas y/o críticas al contrario) carecieran de influencia en su audiencia, el hecho de que alguien se convierta en seguidor (en Twitter) o en fan (en Facebook), demuestra una simpatía, un grado de querencia hacia una opción política determinada, y de esta manera convertir un clic en la Red en una tendencia fiable de voto.

Willians y Gulati llevaron a cabo varias investigaciones centradas en el análisis de Facebook en la campaña electoral de las elecciones de 2008 en EEUU, y en las que concluyeron que el número de fans constituye un indicador de la viabilidad del candidato y por tanto, del éxito electoral¹¹²: *“Thus while Facebook supporter numbers would not be a useful predictor that foreshadows electoral*

¹¹² Otro de sus trabajos es “The political impact of facebook: Evidence from the 2006 midterm elections and 2008 nomination contest”. (2008). *Politics & Technology Review*, 1:11–21

victory or defeat, the most electable candidates do have more Facebook supporters.”
(Willian y Gulati, 2009, p.4)

En sus análisis incorporan una serie de variables para cada candidato, como su posición de peso en el partido, el hecho de que acudan a su reelección, que se encuentren en una campaña competitiva y las características de su electorado (alto nivel de educación, blancos, residentes en núcleos urbanos y jóvenes, son más propensos a participar en Facebook).

En “What is a Social Network Worth? Facebook and Vote Share in the 2008 Presidential Primaries”, utilizaron las siguientes variables de análisis:

- El porcentaje de *supporters* en Facebook para cada candidato por estado (capturado uno o dos días antes de la jornada electoral)¹¹³.
- Cuatro variables de control que previamente habían sido utilizadas por otros investigadores (Adkins & Dowdle, 2000-2001; Mayer, 2004; Steger, 2000; Steger, et al., 2004)¹¹⁴.
- Seis variables de control que consideraron relevantes en el contexto del año 2008¹¹⁵.

¹¹³ Hillary y Obama por el Partido Demócrata y McCain, R. Paul, Huckabee y Romney por el Partido Republicano.

¹¹⁴ *Relative spending state estimate, Relative # of events held in state, Relative national media Coverage, Favorite Son.*

¹¹⁵ *Contest type: Caucus, Contest type:*

% Liberals in state, comunes para ambas opciones

% African Americans in state, % Hispanics in state y % College grads in state, para los demócratas

% Catholics in state, % Mormons in state y % Evangelicals in state, para los republicanos

Tabla 4. Tablas de variables de análisis empleadas para análisis de campañas en Facebook, extraídas del trabajo de Willian y Gulati (2008)

Table 1			Table 2				
Multivariate OLS Regression Analysis of Vote Share in 2008 Democratic Presidential Nomination Contests			Multivariate OLS Regression Analysis of Vote Share in 2008 Republican Presidential Nomination Contests				
Independent Variables	Obama	Clinton	Independent Variables	McCain	Paul	Huckabee	Romney
Relative Facebook strength in state	0.375 ***	1.068 ***	Relative Facebook strength in state	0.079	0.371 **	1.064 ***	-0.041
	0.094	0.196		0.236	0.157	0.235	0.260
Relative spending, state estimates	0.539 **	0.324	Relative spending, state estimates	0.652 ***	-0.097	1.273 ***	0.423
	0.262	0.280		0.095	0.188	0.329	0.282
Relative # of events held in state	0.066 *	0.106 ***	Relative # of events held in state	0.055	0.017	0.192 ***	0.139 **
	0.036	0.033		0.045	0.039	0.050	0.054
Relative national media Coverage	-0.170 *	-0.023	Relative national media Coverage	-0.346 **	-1.278	-0.228	-0.197
	0.099	0.064		0.156	3.982	0.201	0.194
Favorite Son	9.249	-6.020	Favorite Son	10.060	-4.193	-22.374	23.799 ***
	5.948	6.245		10.006	6.129	10.057	8.162
Contest type: Caucus	14.490 ***	-11.669 ***	Contest type: Caucus	-8.791 **	6.562	1.769	10.966 ***
	2.469	1.903		3.850	2.650	2.714	3.674
Contest type: Closed primary	-1.941	2.101 ***	Contest type: Closed primary	3.757	1.411	-1.347	3.894
	2.270	1.877		3.254	2.041	2.529	3.605
% liberals in state	-0.435 ***	0.310 ***	% liberals in state	0.443 **	-0.090	-0.106	-0.077
	0.128	0.098		0.215	0.132	0.162	0.305
% African Americans in state	0.458 ***	-0.563 ***	% Catholics in state	0.591 **	-0.110	0.260	-0.996 ***
	0.092	0.075		0.268	0.164	0.196	0.322
% Hispanics in state	-0.149	0.136	% Mormons in state	0.004	-0.029	0.034	0.249
	0.101	0.084		0.247	0.172	0.203	0.348
% College grads in state	1.486 ***	-1.001 ***	% Evangelicals in state	0.296	-0.102	0.372 *	-0.994 ***
	0.277	0.217		0.270	0.176	0.209	0.291
Intercept	-21.484 *	24.704 *	Intercept	-21.759	4.768	-27.759	63.139
	12.788	13.336		19.464	15.524	17.280	26.001
N	54	54	N	53	52	41	29
R ²	0.765	0.805	R ²	0.859	0.182	0.838	0.842
Note: Bold entries are unstandardized regression coefficients; standard errors are in italics. * p < .10, ** p < .05, *** p < .01.			Note: Bold entries are unstandardized regression coefficients; standard errors are in italics. * p < .10, ** p < .05, *** p < .01.				

Sus resultados, en lo que a Facebook se refiere, afirman que cuanto más se incrementa el número de seguidores de cualquiera de los candidatos en esta red en relación con otros candidatos en el mismo estado, mejor resultado electoral obtiene en el mismo, y que cada porcentaje de incremento supone un aumento en el porcentaje votos finales recabados en dicho estado.

En sus conclusiones, aunque admiten que el valor del “facebook support” no puede tomarse en todos los casos como un elemento de predicción único y fiable del resultado electoral, reconocen que el número de seguidores puede reflejar el efecto de algún recurso de peso en la campaña que no se ha medido: los miembros de una comunidad muestran su entusiasmo y preferencia de voto hacia una opción política haciendo público su “like” en la página del candidato, y el número de fans o

supporters es un indicador de la eficacia con la que un candidato está organizando y movilizando a sus bases.

“Our major conclusion is that online campaign activity is an important additional indicator of candidate viability that is independent of traditional measures like expenditures, media coverage and organizing activities, represented in our model by campaign events.” (Willians y Gulati, 2008, p.16)

En un estudio posterior, “Explaining Facebook Support in the 2008 Congressional Election Cycle”, los autores investigaron qué explicaría el grado de soporte en Facebook y qué evaluaciones estratégicas de campaña se están haciendo sobre esta comunidad de seguidores. Sus principales conclusiones fueron las siguientes (Willians y Gulati, 2009, p.2):

- La cantidad de contribuciones conseguida por una campaña es la variable más importante a la hora de explicar el grado de soporte en facebook, especialmente para candidatos que se presentan a su reelección.
- Los Demócratas tienen más seguidores que los Republicanos, y los candidatos que se presentan a su reelección más que los que compiten con ellos (*challengers*) o aquellos que no compiten con un candidato que se presenta de nuevo (*open seats*).
- Una campaña competitiva incrementa el número de seguidores para los candidatos que compiten con el candidato a reelección.
- En los distritos con un alto porcentaje de ciudadanos con educación universitaria, el número de seguidores es mayor, pero disminuye con un alto porcentaje de jóvenes.
- Los candidatos que mantienen activas sus páginas en Facebook, consiguen mayor número de seguidores que aquellos que apenas la actualizan.

El estudio, basado en la recolección de los datos de las páginas Facebook de aproximadamente 184 candidatos, identificó ocho variables independientes para la variable dependiente “número de seguidores en Facebook”, cuatro relacionadas con características electorales (challengers, open seats, contribuciones recibidas, competitividad de la campaña) y cuatro indicadores de la circunscripción (porcentaje de ciudadanos de más de 24 años con título universitario, porcentaje de ciudadanos clasificados como “blancos”, porcentaje de residentes menores de 14 años y porcentaje de residentes en áreas rurales).

Tabla 5. Tabla de variables de análisis empleadas para análisis de campañas en Facebook, extraídas del trabajo de Willian y Gulati (2009)

Table 1				
OLS Regression Analysis of Facebook Supporters in in the 2008 House Races				
Independent Variables	Initial Model		Final Model	
<i>Party (Republicans=reference category)</i>				
Democrats	121.478 ***	0.121	120.135 ***	0.119
	38.359		37.420	
<i>Incumbency Status (Incumbents=reference category)</i>				
Challengers	-166.768 ***	-0.151	-165.199 ***	-0.150
	47.087		46.556	
Open seat candidates	-169.193 **	-0.096	-166.929 **	-0.095
	70.514		69.874	
Contributions received	0.000 ***	0.300	0.000 ***	0.300
	0.000		0.000	
Competitive seat	-91.106 *	-0.084	-90.308 *	-0.083
	49.325		49.172	
Percent w/college degrees	7.993 ***	0.147	6.294 ***	0.115
	2.73		2.132	
Percent under age 18	-30.273 ***	-0.173	-30.632 ***	-0.175
	7.324		6.725	
Percent white	-0.634	-0.074		
	1.213			
Percent rural	1.411	0.016		
	1.416			
Intercept	761.595 ***		797.563 ***	
	247.671		196.968	
N	588		588	
Adjusted R ²	0.204		0.205	
Standard error	446.135		445.749	
Note: Bold entries are unstandardized coefficients; standard errors are in italics.				
* p < .05, ** p < .01, *** p < .001				

Sus resultados explican que el hecho de que así como el dinero con que cuenta un candidato para su campaña electoral incrementa la visibilidad en los medios de comunicación tradicionales, es de esperar que este reflejo ocurra de

similar modo en Facebook. Es por esto también, que los candidatos que se presentan a su reelección, con una campaña que favorece la captación de fondos, mayor cobertura ofrecida por los medios y mayor reconocimiento personal, consigan mayor número de seguidores que sus oponentes, de igual modo que han conseguido un mayor ratio de reelección en las citadas elecciones de 2008. Por otro lado, sus resultados “negativos” acerca de los distritos con mayor número de jóvenes, estriba en que el número de votantes a estas edades es más bien reducido. Finalmente, el porcentaje de seguidores en Facebook es mayor para aquellos candidatos que mantienen activas sus páginas que para aquellos que no muestran actividad, siendo las actividades que requieren bajo esfuerzo de los usuarios las más premiadas (*Wall post o campaign videos*), y por el contrario, penalizadas aquellas que demandan un esfuerzo mayor (*Fan Photo activity*).

Tabla 6. Comparación del número de seguidores en Facebook según actividad de candidatos, extraída del trabajo de de Willian y Gulati (2009)

Table 3 Comparison of Numer of Facebook Supporters for Candidates With				
Updating Activity	Average # Updated	Average # Not Updated	Significance	R2 Supporters & Updating
Wallposts	365.95	83.57 *		0.843
N	131	7 ¹		138.000
Fan Photos	355.87	339.88		0.264
N	85	51		136
Campaign videos	357.26	231.18 *		0.125
N	43	57		100
Campaign events	341	244.42		0.126
N	35	65		100
*p < .05, **p < .01, ***p < .001				
¹ Substituting the Mean cutpoint the values are 659.42 vs. 175.40 (p < .001); substituting the Median cupoint, the values are 548.49 vs. 147.24 (p < .001)				

Una vez más y coincidiendo con las apreciaciones de investigadores, como el ya citado Daniel Gayo-Avello (Gayo-Avello, 2011), las dificultades para el desarrollo o análisis de una campaña electoral en la red parecen encontrar significación en la constatación de que todavía no suponen un espacio de información y opinión para una muestra suficientemente representativa de la sociedad y/o electores:

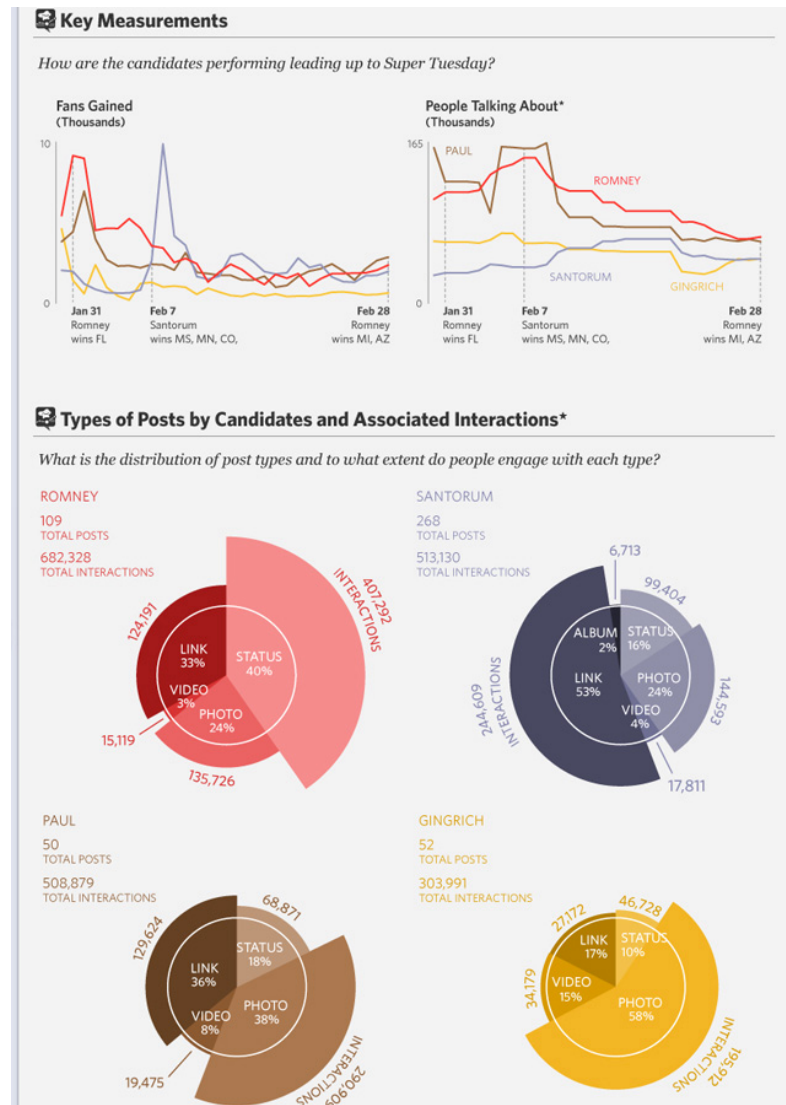
“In a younger district, Facebook would have played a larger role. For this election, we could not afford to spend too much time beefing up Facebook because press releases and other methods are more important and effective ways of campaigning in this district.” Melissa, campaign staff of Christine Jennings (D-FL). (Willians y Gulati, 2009, p.10)

Hoy en día las herramientas de captura y análisis y representación de los datos vertidos en las redes sociales han mejorado sensiblemente y son capaces de recabar la información y tratarla en un breve espacio de tiempo y de forma prácticamente automática. Monitorizando la actividad de las páginas Fan de los candidatos, algunas empresas de consultoría en los social media como Social Bakers ya realiza el seguimiento en Facebook de los candidatos a las elecciones de 2012 en EEUU, permitiendo consultar los datos y exportarlos en diferentes formatos¹¹⁶.

¹¹⁶ <http://www.socialbakers.com/elections>. (Consultado el 03/04/2012)

En España, aunque su herramienta no establece análisis comparativos entre opciones políticas, la herramienta de reciente aparición Smetrica.com permite obtener los datos de distintas páginas de candidatos y partidos políticos en España durante un periodo de tiempo fijado a voluntad.

Ilustración 11. Ejemplo del análisis Facebook de la empresa Social Backers en el seguimiento de las elecciones 2012 en EEUU



3.3.2. Google Trends como herramienta de predicción electoral

Google Trends es una herramienta de Google que muestra los términos de búsqueda más populares del pasado reciente presentando una lista con los términos más buscados en ese momento, a la vez que permite preguntar por la evolución de

cualquier concepto o término de búsqueda. También permite incluir varios términos y recibir en respuesta una gráfica que incluye la frecuencia con que cada término ha sido objeto de búsqueda en Google durante un periodo de tiempo determinado a voluntad¹¹⁷. Dado que Google es el motor de búsqueda que domina en Internet, siendo utilizado por la gran mayoría de la población que navega en la Red, los resultados que ofrece son tomados como las mejores respuestas para la información que estamos buscando. De hecho, los desarrolladores de páginas web trabajan teniendo en cuenta las particularidades de su algoritmo de búsqueda en aras de que Google presente sus páginas Web entre las primeras que resulten de una búsqueda por un concepto determinado. Siendo así, cabría conjeturar que si se realiza una consulta en esta herramienta por los candidatos que toman parte en una campaña electoral, obtendremos la cantidad de veces que han sido demandados por la población durante la campaña electoral, lo cual podría traducirse en que a mayor número de búsquedas de un candidato este tendrá mayor posibilidades de éxito electoral que sus oponentes.

En los trabajos mencionados anteriormente en este marco teórico (Ginsberg et al., 2009, y Choi y Varian, 2009) sus métodos se basaban en la capacidad de Google Trend para saber cómo un término de búsqueda (un producto, una marca, un personaje, etc.) ha evolucionado a lo largo del tiempo en función de la cantidad de veces que ha sido buscado en un periodo determinado. Pero Google Trend también está preparado para comparar las frecuencias de búsqueda de varios términos, particularidad que ha permitido a Lui, Metaxas y Mustafaraj (2011) realizar un análisis sobre *The 2008 and 2010 US Congressional Elections*, comparando la popularidad resulta en las búsquedas de pares de candidatos compitiendo por una posición política¹¹⁸.

¹¹⁷ Permite además, a través de Google Insights for Search, <http://www.google.com/insights/search/>, exportar los resultados detallados en un fichero con formato CSV, que puede ser importado y tratado con una herramienta básica como Excel.

¹¹⁸ Sus resultados están accesibles en una gráfica interactiva en

Ilustración 12. Ejemplo de la gráfica interactiva de la visualización de los resultados del uso de Google Trends como herramienta de predicción electoral, de Lui, Metaxas y Mustafaraj (2011)



Las líneas básicas de su metodología fueron las siguientes:

- Recolectar los datos de Google Trends para ambos procesos, centrándose en la semana previa a la celebración de las elecciones y utilizando como términos de búsqueda los nombres de los candidatos tal y cómo los difunde el diario *New York Times*¹¹⁹.

<http://cs.wellesley.edu/~webtrust/insights/>

¹¹⁹ <http://www.nytimes.com/politics/>

No disponen de las encuestas de 2008 ya que fueron borradas tras la elección.

- Establecer una comparación de estos con las encuestas publicadas por el *New York Times* y con el ratio de “*incumbency re-election*” de ambas elecciones¹²⁰
- Marcar una segunda línea de análisis comparado en la que sólo incluyen las disputas (*racas*) o pares de candidatos con una alta presencia en los medios de comunicación a lo largo del mes previo a las elecciones (*highly contested*).
- Separar las “*racas*” en las que alguno de los candidatos Google Trends no había registrado dato alguno de búsqueda (grupo T1), de aquellas en las que ambos candidatos si obtuvieron registros (grupo T2).

De forma somera, sus resultados fueron negativos para la predicción con esta herramienta salvo en las elecciones de 2008 en las que se tomaron en cuenta aquellas contiendas en las que sólo uno de los candidatos mostró visibilidad en la herramienta de búsqueda.

Ilustración 13. Figura 1 de resultados de búsquedas en Google Trends (para grupos T1 y T2) de Lui, Metaxas y Mustafaraj (2011)

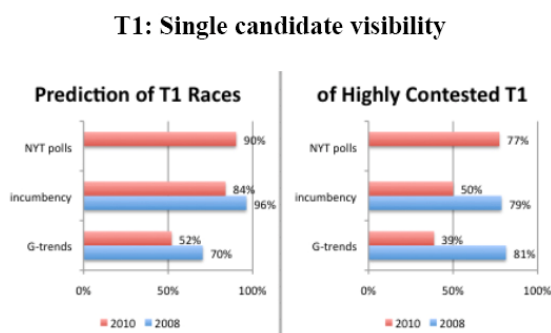


FIGURE 4: G-trends did not compare well with both incumbency and NYT predictions for the group T1 that included races with a single candidate having no visibility. It did quite well for the highly contested races in 2008, but quite poorly in 2010.

¹²⁰ Dado que en los EEUU el porcentaje de victorias en candidatos que se presentan a su re-elección en extremadamente alto, decidieron tomar el ratio de *incumbency* de una elección como predictor electoral.

Ilustración 14. Figura 2 de resultados de búsquedas en Google Trends (para grupos T1 y T2) de Lui, Metaxas y Mustafaraj (2011)

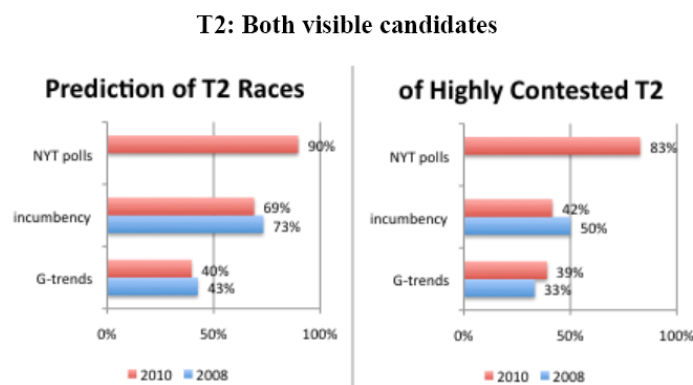


FIGURE 5: For the most competitive group T2 for which both candidates had visibility, G-trends did not compare well with incumbency. In the all-important high contested races, its performance was ever worse. The polls are proven to be the best way to make good predictions.

Los autores argumentan que las causas de estas desviaciones pueden deberse por un lado al hecho de que cuando un usuario busca por un candidato y no por su oponente, el resultado es que este último no alcanza los niveles mínimos de curiosidad y/o visibilidad (permaneciendo en el vacío) lo que repercute más positivamente en el candidato buscado; por otro, que los electores pueden buscar por un candidato pero con el objetivo de encontrar información negativa lo que provoca ese margen de error con los resultados electorales; y finalmente, que aquellos candidatos que ya ocupan un cargo político sobresaliente pueden ser buscados a causa de esta actividad y no tratando de encontrar información sobre su campaña:

“A variable that may have affected G-trends effectiveness as a tool for predicting political elections is the sentiment of a user’s query. It is difficult, though not impossible, to determine the circumstances behind a user’s search of the profile of a certain candidate to make a guess about that candidate’s public image and why a user might be interested in the candidate. This is part of future research that we plan for the next stage of our work.”
(Lui, Metaxas y Mustafaraj, 2011, p.7)

Hemos visto que así como Google Trends ha resultado una herramienta de utilidad en el campo de la medicina, concretamente en el de la vigilancia de epidemias, su capacidad para coincidir con resultados electorales, mediante la búsqueda de los nombres de los candidatos, no ha obtenido los buenos ratios de acierto que pudieran esperarse. A los diversos motivos que ya han sido expuestos por los autores de la investigación, podría sumarse el hecho de que los usuarios no tienen por qué buscar exactamente a un candidato por el nombre de referencia empleado en un medio de comunicación, mostrándose quizás como apropiado proponer para cada candidato, la suma de todas las posibles búsquedas que sobre él pudieran hacerse introduciendo otras “denominaciones”. No obstante, serían necesarios nuevos estudios en este campo para determinar el grado de acierto de esta herramienta en relación con su aproximación a los resultados electorales, puesto que la fiabilidad de las técnicas de Google es aval de su profesionalidad y junto con su uso extensivo en todo el mundo, la convierten en una opción nada desdeñable.

3.3.3. Twitter como herramienta de predicción electoral

Debido a los positivos resultados obtenidos en primeros estudios sobre la capacidad del análisis de diferentes redes sociales como medios para obtener de forma más o menos fácil e instantánea el sentimiento global o las tendencias de consumo de la población a tenor de lo que la mayoría de la misma difunde en Internet, se ha apelado a la idea de que lo que la gente está publicando en su blog, o buscando en google o difundiendo en sus perfiles de Facebook y/o Twitter, proporciona un reflejo del sentimiento colectivo, lo cual puede traducirse en tendencia hacia un punto determinado. En este sentido, recientes estudios han anunciado que es posible predecir un resultado electoral mediante el análisis de este contenido social.

Concretamente, en “Predicting elections with Twitter: What 140 characters reveal about political sentiment”, se concluye que en las elecciones de 2009 en Alemania, Twitter se revela como predictor del resultado electoral: *“We find that the mere number of messages mentioning a party reflects the election results. Moreover,*

joint mentions of two parties are in line with real world political ties and coalitions.” (Tumasjan et al., 2010, p.178) Su metodología de análisis se basa en estudios anteriores y similares, como el de Véronis referido a las elecciones presidenciales de 2007 en Francia, en el que se expresa que el número de apariciones o de menciones de un candidato en la prensa puede ser un buen predictor del resultado de la contienda¹²¹.

En su estudio, en el que se reclutaron 104.003 *tweets* publicados en las semanas anteriores a las elecciones al parlamento alemán del 27 de septiembre de 2009, su investigación trataba de dar respuesta a las siguientes preguntas:

- ¿Es Twitter una plataforma de deliberación política online?
- ¿Hasta qué punto puede Twitter informarnos sobre el “sentimiento” del electorado?
- ¿Puede Twitter servir como predictor electoral?

Es la última pregunta la que interesa al objeto de nuestra investigación y en la que vamos a centrarnos: ¿Puede emplearse Twitter como herramienta de predicción del resultado electoral, monitorizando una campaña cualquiera en ella? Y de ser así, ¿puede el análisis de una campaña electoral en esta Red Social sustituir o, al menos convivir, con los métodos de encuestas tradicionales (encuesta personal, encuesta telefónica, encuesta postal, *focus group*) empleados a día de hoy?

Tumasjan et al., (2010) compararon, por un lado, la cuota de atención (seguidores, menciones, replies, RT, hashtag, etc.)¹²² que los partidos políticos¹²³

¹²¹ VÉRONIS, Jean. (2007). “Citations dans la presse et résultats du premier tour de la présidentielle” Internet. <http://aixtal.blogspot.com/2007/04/2007-la-presse-fait-mieux-que-les.html>. (Consultado el 15/12/2009)

¹²² Los autores se refieren a “*the share of tweets*” y “*all mentions*” pero no detallan qué elementos particulares de la conversación en Twitter han tomando en consideración, ya que existen diferentes formas de mencionar a un partido o candidato en esta red.

¹²³ Por contexto y citas, se extrae que su análisis se refiere a menciones de partidos únicamente y no toman en cuenta las menciones a candidatos.

recibieron en Twitter con los resultados electorales definitivos, y por otro, analizaron cuando los tweets pueden mostrarnos las cadenas o enlaces ideológicos entre los partidos y el potencial de coaliciones políticas que se formarían pasadas las elecciones.

Con el fin de comparar los resultados de Twitter con los resultados de encuestas tradicionales, los autores comparan los márgenes de error de sus resultados en Twitter con los presentados por las últimas encuestas electorales difundidas en Alemania. Su conclusión, a la vista del análisis comparativo llevado a cabo, es la siguiente: *“In conclusion, the mere number of tweets mentioning a political party can be considered a plausible reflection of the vote share and its predictive power even comes close to traditional election polls.”* (Tumasjan et al., 2010, p.183)

Tabla 7. Tablas de resultados del análisis de resultados elecciones alemanas de 2009 según Twitter, extraída del trabajo de Tumasjan et al., (2010)

Table 4: Share of tweets and election results

Party	All mentions		Election	
	Number of tweets	Share of Twitter traffic	Election result*	Prediction error
CDU	30,886	30.1%	29.0%	1.0%
CSU	5,748	5.6%	6.9%	1.3%
SPD	27,356	26.6%	24.5%	2.2%
FDP	17,737	17.3%	15.5%	1.7%
LINKE	12,689	12.4%	12.7%	0.3%
Grüne	8,250	8.0%	11.4%	3.3%
			MAE:	1.65%

* Adjusted to reflect only the 6 main parties in our sample

Table 5: Forecast accuracy of various election polls

Source	MAE (last poll)
Twitter	1.65%
Forsa	0.84%
Allensbach	0.80%
Emnid	1.04%
Forschungsgruppe Wahlen	1.48%
GMS	1.40%
Infratest/dimap	1.28%

Un método similar pero referido al número de seguidores o *followers* de un perfil en Twitter, fue mantenido por Dan Zarrella (2010), científico y estudioso de los social media. Con motivo de las elecciones legislativas 2010 en EEUU, el autor eligió al azar treinta disputas electorales con el fin de dar respuesta a la pregunta “Can twitter predict elections?”. Su método se basaba en comparar el número de seguidores conseguido por cada candidato en Twitter con las predicciones electorales

que estos habían obtenido y que se recogían en *Real Clear Politics*¹²⁴. Concluyó que en el 71% de los casos, el candidato con mayor número de seguidores era también el candidato que ocupaba la primera posición en las encuestas.

En este momento podríamos ya preguntarnos: ¿Cuál es el valor de un *follower*? ¿Qué actividad de las que pueden realizarse en Twitter confiere más peso o capacidad de influencia al usuario al que se refieren? ¿Tienen las menciones, los seguidores, *retweets* (RT's) y conversaciones (Replies) el mismo valor en un análisis realizado con los datos obtenidos en esta Red?

En “Measuring user influence in Twitter: The million follower fallacy”, Cha et al. (2010) afirman que el número de seguidores (denominado *indegree*) es un signo que revela la popularidad de un usuario (fuentes de noticias y figuras públicas) pero que no está relacionado con su capacidad de influencia entre la audiencia del mismo modo en que sí lo están menciones y *retweets*: “*users who have high indegree do not necessarily spawn many retweets or mentions. This finding suggests that indegree alone reveals very little about the influence of a user.*” (Cha et al., 2010, p. 8)

Su estudio viene a reforzar la hipótesis de la “falacia del millón de seguidores” refrendada en un anterior trabajo de investigación soportado por A. Avnit¹²⁵ (2009) en el que se evidenció que algunos usuarios siguen a otros simplemente por educación, convirtiéndose en *followers* de aquellos que también les siguen a ellos pero que apenas leen los tweets publicados por estos usuarios a los que siguen.

¹²⁴ Utilizó las cuentas Twitter que se publicaban en las páginas Web oficiales de cada candidato.

http://www.realclearpolitics.com/epolls/latest_polls/elections/. Presenta una media de las últimas encuestas obtenidas de múltiples fuentes.

¹²⁵ AVNIT, Adi. (2009). “The Million Followers Fallacy” *Internet Draft, Pravda Media*. <http://tinyurl.com/nshcjg> (Consultado el 21/02/2012)

La metodología de su investigación, en la que analizaron 2.000 millones de vínculos ende seguimientos entre 54 millones de usuarios que produjeron un total de 1.700 mil millones de tweets, consitió en analizar por separado tres indicadores de influencia de un usuario en twitter: Indegree o número de seguidores (indicador del tamaño de su audiencia), el número de retweets conseguido por los usuarios (indicador de la habilidad de un usuario de generar contenido con valor suficiente para ser “pasado”) y el número de veces en que éste es mencionado (indicador de la habilidad de un usuario para atraer a su audiencia). (Cha et al., 2010, p. 3)

Algunas de sus conclusiones más significativas son las siguientes (Cha et al., 2010, p. 4):

- Los usuarios con mayor número de seguidores fueron los medios de información o generadores de noticias y las figuras públicas.
- Los usuarios más “retuiteados” (aquellos que obtuvieron mayor número de *retweets*) fueron los servicios de agregación de contenido (ej. Mashable), hombres de negocios (ej. Guy Kawasaki) y sitios de noticias (ej. The New York times).
- Los usuarios más mencionados fueron las *celebrities*.

La red de microblogging Twitter se ofrece, al igual que otras redes en las que sus usuarios emiten en forma de texto sus pensamientos, sentimientos y opiniones acerca de una gran variedad de asuntos, como un espacio del que es posible extraer la opinión pública utilizando modelos de programación lingüística. Estos métodos de análisis se conocen como *sentiment* o *sentimental analysis*. La pregunta que aquí trataremos de resolver es: ¿podemos analizar los datos públicos disponibles en estas redes para captar las actitudes y tendencias de voto de la población/electores del mismo modo que las encuestas de opinión pública vienen haciendo hasta ahora?

El *sentimental analysis* relacionado con la expresión de una opinión política o ideológica ha sido también objeto de interés en estudios que no emplearon Twitter como fuente de datos. En “Exploring the Characteristics of Opinion Expressions for

Political Opinion Classification”, Yu et al., (2008), reportaron los resultados de una serie de experimentos -tres- diseñados para explorar las características que podrían definir las formas de expresión que las personas emplean que nos permitan inferir su opinión política. Trataron de establecer una comparación entre las expresiones lanzadas por los senadores y congresistas americanos en sus intervenciones en ambas Cámaras en el periodo comprendido entre 1989 a 2006 con las expresiones que se utilizan en el entorno empresarial, concretamente en artículos de noticias y en artículos de opinión sobre películas cinematográficas. Para el análisis lingüístico del sentimiento, sus herramientas fueron dos: el LIWC (*Linguistic Inquiry and Word Count*) y el SVMs (*Support Vector Machines*). Como fuente para capturar expresiones de los políticos, descargaron los discursos de los senadores (*Thomas.gov*) y de los congresistas (*2005 House speeches*) en sendas bases de datos, que garantizaban su autenticidad¹²⁶. Y como origen de las noticias sobre el contexto empresarial, 130.000 artículos publicados en *Wal-Mart*¹²⁷. Encontraron que el promedio de nivel “sentimental” en política es bastante bajo, mayor que el reportado como neutro en los artículos noticiosos pero mucho más bajo que el encontrado en los artículos de críticas de películas. Y aunque en estudios previos¹²⁸ los adjetivos habían sido tomados como indicadores significativos del sentimiento vertido en una expresión, hallaron que los adjetivos (incluidos los de forma afectiva) no constituyen un indicador de peso en el análisis de sentimiento político. En vez de eso, los nombres que hacía referencia a los temas tratados, reflejados como neutros en este tipo de análisis sentimental, constituían un elemento de primer orden en sí mismos.

¹²⁶ THOMAS, Matt, PANG, Bo, y LEE, Lillian. (2006). “Get out the vote: Determining support or opposition from Congressional floordebate transcripts”. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP2006)*, 327-335

¹²⁷ PANG, Bo, LEE, Lillian y VAITHYANATHAN, Shivakumar. (2002). “Thumps up?: Sentiment classification using machine learning techniques.” *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, Volume 10*, 79-86

Actual Walmart desde 2008. Cadena internacional de supermercados de nacionalidad americana.

¹²⁸ HATZIVASSILOGLU, Vasileios y MCKEOWN, Kathleen R. (1997) “Predicting the semantic orientation of adjectives”. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics (ACL '98)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 174-181

Concluyen que con sus experimentos no lograron obtener un método satisfactorio para la clasificación sentimental de las opiniones políticas, sugiriendo que encontrar nuevas vías de aproximación al tema, aportaría un gran valor a investigaciones futuras sobre métodos de clasificación lingüística en opiniones políticas:

“Once we have properly identified a person’s ideology, we may be able to predict his or her opinions on various political issues. It is our goal for future work to explore viable approaches for ideology based on political opinion classification.” (Yu et al., 2008, p.90)

Si bien no se refiere a Twitter, los resultados de esta investigación comienzan a indicarnos que el *sentiment analysis* es una disciplina cuanto menos difícil y a la que es necesaria contribuir con programas de análisis del lenguaje avanzadas. En las siguientes referencias sobre el tema que mencionaremos en los párrafos siguientes, volveremos a constatar que existen dificultades, de índole variada, para que este tipo de análisis encuentre una base firme en la que apoyar su confianza.¹²⁹

En “Form tweets to polls: linking text sentiment to public opinion time series”, O’Connor et al. (2010) relacionaron la opinion pública en EEUU en el periodo 2008-2009 medida en las encuestas tradicionales con el sentimiento reflejado en los tweets (*sentiment analysis*). Sus resultados eran estrechamente coincidentes en el caso de “confianza del consumidor” (con una correlación del 80%) y con diferente resultado en los temas de “opinión política”: para *presidential job approval* en 2009 los datos obtenidos de Twitter replican los obtenidos en las encuestas y para *presidential elections polls* en 2008 la correlación no fue tan significativa: *“While our results vary across datasets, in several cases the correlations are as high as 80%, and capture important large-scale trends. The results highlight the potencial of*

¹²⁹ Sin ánimo de entrar en el extenso campo de la lingüística, si conviene apuntar aquí que la diferencia de giros y expresiones existente entre las diferentes lenguas conocidas, constituye un reto y que además, los adelantos, en forma de herramientas, configuradas para una lengua determinada, resultan inadecuadas para una lengua y/o entorno diferente.

text streams as a substitute and supplement for traditional polling.” (O’Connor, Rammath, Bryen y Noah, 2010, p.1)

Los autores recolectaron mil millones de mensajes publicados en Twitter durante los años 2008 y 2009 capturados mediante el API de libre disposición que ofrece dicha red, filtrando dichos mensajes por la presencia o no de una serie de keywords¹³⁰ decididas manualmente, y definiendo la positividad o negatividad de los mensajes siguiendo el patrón de palabras positivas, negativas o neutras recogido en el thesaurus *OpinionFinder* (los autores han eludido las distinciones entre palabras fuerte o débiles). Las encuestas con las que establecieron comparaciones respectivas fueron para la “confianza del consumidor”, el *Index of Consumer Sentiment* (ICS) de Reuters y la Universidad de Michigan, y el *Economic Confidence Index* de la consultora Gallup, para *presidential job approval* el tracking diario de la consultora Gallup, y para *presidential elections polls* la compilación de encuestas recogidas por Pollster.com.

Concluyen que “*while the results do not come without caution, it is encouraging that expensive and time-intensive polling can be supplemented or supplanted with the simple-to-gather text data that is generated from on line social networking.*” (O’Connor et al., 2010, p.8) y argumentan que las diferencias encontradas en las comparaciones establecidas en su análisis entre los datos recolectados en Twitter y los recogidos en las encuestas, podrían deberse a las siguientes razones:

- Las diferencias evolutivas temporales se sustentarían en que en el año 2008 la red Twitter contaba con menos usuarios que en 2009, pudiendo no representar a la población en general.

¹³⁰ Para “confianza del consumidor”, *economy, job y jobs*, para la aprobación de la gestión del presidente, *obama*, y para elecciones, *obama y mccain*.

- Los mensajes que mencionan a Obama en 2008 son mayores que los recibidos para McCain debido a que son un reflejo o indicador de las noticias y eventos electorales, más favorables o predispuestos a Obama.
- Son necesarias herramientas más avanzadas de NLP (programación neurolingüística) para mejorar dichas estimaciones.

La dificultad para realizar estudios de “sentiment analysis” fue también constatada por Asur y Huberman (2010) en su estudio “Predicting the future within social media”: *“Sentiment analysis is a well-studied problem in linguistics and machine learning, with different classifiers and language models employed in earlier work [13], [14]. It is common to express this as a classification problem where a given text needs to be labeled as Positive, Negative or Neutral.”* (Asur y Huberman 2010, p.6) Los autores recurrieron a diversas herramientas y diccionarios para proseguir con su investigación, concluyendo, como ya anunciamos anteriormente, que el análisis sentimental no les procuro diferencia o conocimiento substancial sobre el análisis exhaustivo de *tweets*.

Recientemente se viene expresando un creciente interés en construir una propuesta de clasificadores textuales o lingüísticos de la opinión pública capaces de catalogar como favorables o desfavorables las opiniones vertidas por los individuos sobre una propuesta o tema político. Los enfoques metodológicos que se han empleado hasta ahora (diccionario de términos empleado en la búsqueda de palabras y textos pre-etiquetados a modo de ejemplo) no han conseguido resultados tan aproximados como los conseguidos en las críticas periodísticas sobre temas de dominio público como los negocios y los bienes de consumo. Yu et al., (2008) realizaron un experimento en aras de conseguir un patrón metodológico de etiquetado sentimental de los mensajes (positivo, negativo o neutro) que ayudara en la clasificación del análisis del sentimiento de la opinión pública, analizando los discursos del Senado¹³¹ de los EEUU durante el periodo 1989-2006 y los discursos

¹³¹ Conseguidos en la página Thomas.gov.

del Congreso¹³² en el año 2005, comparándolos con los artículos sobre negocios de Wal-Mart¹³³ en 2006 y las críticas sobre películas empleadas por Pang, Lee y Vaithyanathan (2002)¹³⁴. Encontraron que el nivel de sentimiento expresado por los políticos en sus disertaciones es bastante bajo, más alto que en los artículos o noticias (neutro) pero mucho menor que el expresado en las críticas de las películas. Por otro lado y contrariamente a lo que sucede en otros ámbitos, mientras que se venía considerando que los adjetivos afectivos cumplían un papel definitivo en la clasificación de los sentimientos, en el dominio de la política, estas palabras no son los mayores indicadores del tipo de la información que se vierte, siendo los nombres modos de expresión del sentimiento político válidos en sí mismos. Sugieren que conocer de antemano la ideología de las personas analizadas supondría una ventaja y una ayuda a la hora de intentar establecer una metodología de análisis sentimental.

Tabla 8. Tabla de clasificación de sentimientos reflejados en Twitter acerca en los discursos del Senado de EEUU, extraída del trabajo de Yu et al., (2008)

Table 5: Sentiment levels in different domains (%)			
Data set	Sentiment	Positive	Negative
News articles	2.78	1.91	0.84
Senate debate	3.43	2.35	1.05
Movie reviews	4.98	3.04	1.94

Su estudio viene a reafirmar la dificultad que existe hoy en día para llevar a cabo un *sentiment analysis* eficaz, más aún, en el campo de la opinión política.

¹³² Extraídos del estudio de THOMAS, Matt, PANG, Bo, y LEE, Lillian. (2006). "Get out the vote: Determining support or opposition from Congressional floordebate transcripts". *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (EMNLP2006), 327-335.

¹³³ 130.000 artículos sobre productos de consumo.

¹³⁴ PANG, Bo, LEE, Lillian y VAITHYANATHAN, Shivakumar. (2002). "Thumps up?: Sentiment classification using machine learning techniques." *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, Volume 10*, 79-86.

Los estudios mencionados, aplicando técnicas diferentes en distintos momentos sobre los datos capturados en Twitter, han arrojado resultados positivos en lo que se refiere a la capacidad de utilizar esta herramienta como método de predicción electoral en sustitución de las encuestas tradicionales. Gayo-Avello, Metaxas y Mustafaraj (2011) replicaron estas dos metodologías empleadas aplicándolas en un escenario diferente, las elecciones al Congreso de los EEUU en 2010, llegando a la siguiente conclusión:

“Unfortunately, we find no correlation between the analysis results and the electoral outcomes, contradicting previous reports” (Gayo-Avello et al., 2011, p.1).

Aún repitiendo los métodos citados, los autores incluyeron pequeñas variaciones. Respecto al trabajo de Tumasjan et al., (2010) (basado en el volumen de tweets mencionando una opción política y estableciendo la comparación de sus resultados con los resultados electorales), aunque en éste cada tweet que menciona a un partido (candidato) es tomado como un “voto”, en el caso de las elecciones al Congreso de los EEUU no se han contabilizado los tweets en los que se mencionan a candidatos opuestos¹³⁵, y en el caso de O’Connor et al. (2010) (basado en la técnica del *sentiment analysis* y estableciendo la comparación de sus resultados con encuestas pre-electorales), aunque se ha empleado el mismo tesoro, cada tweet podía únicamente pertenecer a una de las tres categorías definidas (positiva, negativa o neutra.), y no a varias de ellas.

El resultado de su investigación arroja un MAE de 17,1% para la comparación del “volumen de tweets” y de 7,6% para el “análisis sentimental”, mientras que se indica que el MAE típico para las encuestas profesionales se encuentra entre un 2-3%. Los autores argumentan que las razones por las que las encuestas profesionales no pueden ser duplicadas por la simple toma de muestras de

¹³⁵ En las elecciones alemanas se trata de 5 partidos mayoritarios, y en estas elecciones EEUU se trata de parejas de candidatos “opuestos” luchando por un mismo sitio.

los datos de las redes sociales, pueden hallarse en el hecho de que no todos los posibles votantes han decidido expresar su opinión en Twitter y por tanto la “muestra” aquí no es representativa del conjunto de los electores, y en que en estas redes los datos que viajan en ellas permiten la manipulación de mano de *spammers* y propagandistas.

No obstante, y aunque los mismos autores expresan en su trabajo que no se pueden obtener resultados diferentes repitiendo el mismo tipo de análisis, su investigación no ha clonado estrictamente los métodos apuntados por anteriores autores, ya que han incluido pequeñas variaciones en cada uno de ellos (citadas anteriormente) y, tanto o más importante, los escenarios y tiempos electorales son bien diferentes. Sin duda y como todos ellos concluyen, son necesarios futuros trabajos en los que las metodologías varíen incluyendo nuevas variables de análisis y perfeccionen sus herramientas de análisis: “... *while according to the state-of-the-art polling techniques, correct predictions require the ability of sampling likely voters randomly and without bias.*” (Gayo-Avello et al., 2011, p.4)

En un estudio posterior realizado en solitario, Daniel Gayo-Avello, vuelve a advertir sobre esta facilidad de convertir los medios sociales en el futuro Literary Digest¹³⁶ tras su estudio retrospectivo dedicado a comparar los datos recabados en Twitter con los resultados electorales durante las Elecciones Presidenciales de EEUU en 2008: “*This paper will attempt to lay devil’s advocate by detailing a study in which such simple approaches largely overestimated Obama’s victory in the 2008 U.S Presidential Elections*” (Gallo-Avello, 2011, p.1). El estudio recogió 250.000 tweets publicados por 20.000 usuarios desde el 1 de junio hasta el 11 de noviembre de 2008,

¹³⁶ La revista Literary Digest llevó a cabo un “sondeo” en relación con el posible resultado de las elecciones presidenciales de 1936. El sondeo mostró que el gobernador republicano de Kansas, Alf Landon, probablemente sería el ganador con una victoria abrumadora. En noviembre, Landon gana sólo Vermont y Maine. El presidente Franklin Delano Roosevelt ganó en los 46 restantes estados de entonces. La revista fue desacreditada por completo a causa de la encuesta y dejó de editarse poco tiempo después. Las técnicas de votación empleadas por la revista fueron las culpables: los encuestados eran en su mayoría lectores de su revista, con un gran nivel adquisitivo y la financiación de contacto de los encuestados se extrajeron de bases de datos de dueños de automóviles y con teléfonos en sus hogares.

geolocalizados en 8 estados representativos de los EEUU (California, Texas, Florida, Indiana, Misouri, Montana, Carolina del Norte y Ohio). Su singularidad estriba en que en vez de “agregar” datos de Twitter, se detectaron las tendencias de voto por cada usuario identificado: “... *to count the number of appereances of a candidate in the user’s tweets assuming the one more frequently mentioned would be the one would vote for*” (Gallo-Avello, 2011, p.7) y utilizando cuatro métodos de análisis sentimental, entre ellos el *semantic orientation*¹³⁷ que se aplicó sólo en aquellos casos en los que el usuario incluía manifiestamente una intención de voto en sus tweets con frases como “I will vote for...”, “I’m not voting...”, etc.

En este campo del “análisis sentimental” encontró diferencias significativas entre la precisión de los resultados para Obama y para McCain, recogiendo este último los valores más pobres y alejados de la realidad, no superando el 18% de precisión, mientras que en el caso del candidato vencedor, los porcentajes superaban el 80% en los cuatro métodos de investigación sentimental empleados. Su primera conclusión es que el análisis sentimental supone un gran desafío y el investigador debe proceder con extrema precaución cuando asume que un “ingenuo” clasificador de palabras/sentimientos puede hacer todo el trabajo de forma más o menos automatizada.

En el campo del “volumen de datos” (Twitter data), el fracaso se hizo también patente, mostrando un MAE global del 13,10% (sumando los resultados individuales de los estados monitorizados) al comparar el porcentaje de “votos Twitter” con el porcentaje de votos conseguidos por Obama. El autor atribuye el error al *self-selection bias* que puede explicarse en el uso que se da de Twitter (Gallo-Avello 2011, p.10):

¹³⁷ TURNEY, Peter D. (2002). “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews”. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 417-424.

- Los votantes que viven en núcleos urbanos así como los jóvenes adultos son más propensos a utilizar Twitter como forma de expresión, y estos mismos muestran una tendencia hacia las opiniones liberales en política.
- Los votantes republicanos hacen un menor uso de Twitter que los demócratas a la hora de expresar sus tendencias políticas. Este hecho se conoce como el “*shy-Republican factor*”.

Ambas afirmaciones son plausibles ya que el estudio de “The internet and the 2008 election” (Smith y Rainie, 2008) manifestaba esta tendencia al desvelar que el 65% de los seguidores online de Obama han consultado información política online comparado 56% de McCain, habiendo aquellos efectuado más contribuciones campaña, más firmas de peticiones online, más comentarios en blogs y más vídeos vistos que el resto del electorado.

Tabla 9. Tablas de análisis sentimental y volumen de datos en Twitter durante las Elecciones Presidenciales de EEUU en 2008, extraídas del estudio de Gayo-Avello (2010)

Method	Precision Obama	Precision McCain	Accuracy
Most frequent candidate	82.4%	7.8%	50.7%
Polarity lexicon	88.8%	17.7%	61.9%
Vote & Flip	92.7%	10.7%	50.6%
Semantic Orientation	92.3%	15.6%	36.7%

Table 3: Performance results for each of the four automatic sentiment analysis methods employed to infer user voting intention.

State	Actual % of Obama votes	% of Twitter "votes"	Twitter error	% of TwitVote "votes"	TwitVote error
California	62.28%	62.70%	0.42%	91.89%	29.61%
Florida	51.42%	66.20%	14.78%	81.32%	29.90%
Indiana	50.50%	64.70%	14.20%	87.88%	37.38%
Missouri	50.07%	68.10%	18.03%	83.61%	33.54%
N. Carolina	50.16%	66.60%	16.44%	98.38%	48.22%
Ohio	52.31%	59.80%	7.49%	86.57%	34.26%
Texas	44.06%	64.40%	20.34%	76.97%	32.91%
		MAE	13.10%	MAE	35.12%

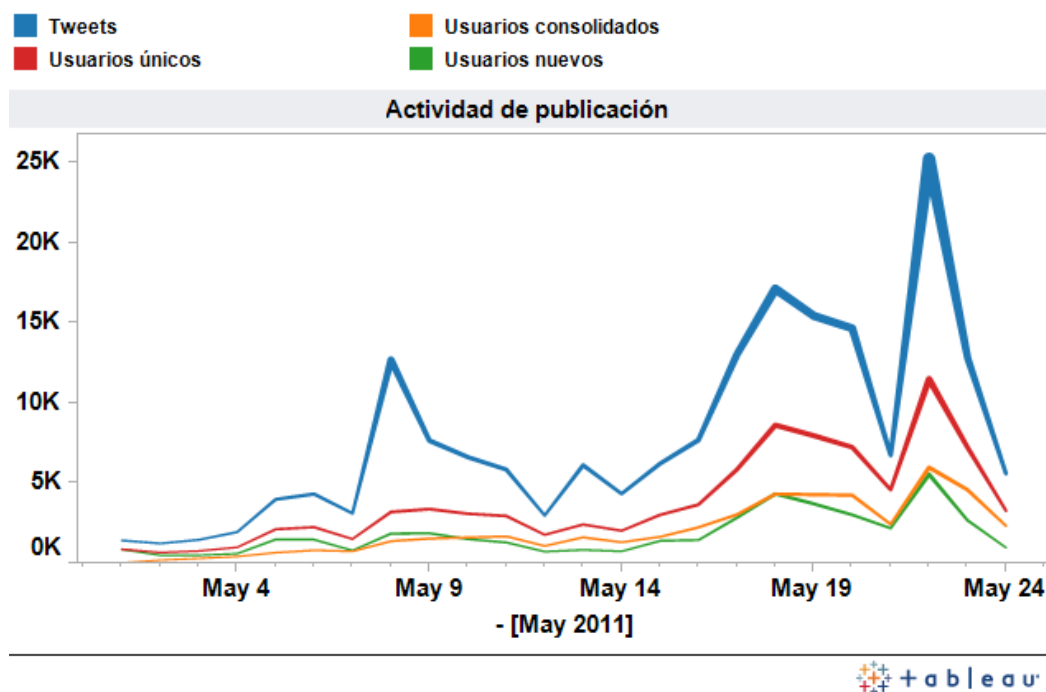
Table 5: Prediction of the 2008 U.S. Presidential Elections according to data collected in Twitter and to the subset of users who issued a vote in TwitVote. The MAE is **very** large and a victory for Obama in Texas is predicted. Nevertheless, the prediction using tweets is **substantially** better than the direct-poll conducted by TwitVote.

En sus conclusiones, Gayo-Avello enumera una serie de lecciones aprendidas y a tener en cuenta (Gallo-Avello, 2011, pp.12-13):

- *“The big data fallacy...”*
- *Watch out for demographic bias...*
- *Beware of naïve sentiment analysis*
- *Silence speaks volumes...*
- *(A few) Past positive results do not guarantee generalization...”*

Por otra parte, uno de sus resultados más “esperables” fue comprobar que la cantidad de informaciones/conversaciones/mensajes que circulan en Twitter crece según progresa el periodo de campaña electoral, mostrando picos acordes con sucesos relevantes que afectan o suceden en la misma, y cuyo volumen decae tras la jornada de votación electoral. Similares efectos han sido comprobados en diversos trabajos de la investigadora Mariluz Congosto de la Universidad Carlos III de Madrid. Concretamente en sus visualizaciones de la evolución temporal de la campaña electoral de las Elecciones Autonómicas de la Comunidad de Madrid 2011, se observa un incremento en la actividad de publicación en Twitter durante el periodo de campaña, que curiosamente desciende en la jornada de reflexión y vuelve a repuntar tras el conocimiento de los resultados electorales. (Fernández y Congosto, 2012)

Ilustración 15. Visualizaciones de la evolución temporal de la campaña electoral de las Elecciones Autonómicas de la Comunidad de Madrid 2011 realizada por Congosto (Fernández y Congosto, 2012)



En este caso, el análisis de redes sociales se revela como un campo nuevo para la investigación en estudios dedicados a analizar la influencia de las campañas electorales en la decisión de voto de los individuos o, al menos, en manifestar que las campañas electorales afectan al individuo que se manifiesta en estos periodos de tiempo más proclive a la participación en el desarrollo y legitimación del sistema democrático. En la introducción del Volumen II de los resultados del trabajo del Grupo de Investigación sobre Campañas electorales (GICE), dirigido por el Dr. Ismael Crespo, centrado en el marco de las Elecciones Generales del año 2000 en España (*La campaña electoral del 2000: partidos, medios de comunicación y electores*) se manifiesta literalmente el papel destacado de los medios de comunicación: “El hecho de que la campaña se manifieste principalmente a través de los medios de comunicación plantea la necesidad de abordar el estudio de los

mismos desde todos los enfoques metodológicos posibles, combinando técnicas de investigación diferente”. (Crespo et al., 2004, p.39)

Desde hace unos años, si bien los medios continúan como proveedores privilegiados de información, inclusive en las campañas electorales, el desarrollo y adopción en masa de Internet como canal de información y comunicación, está modificando el paradigma. Por un lado, los escenarios en los que los líderes políticos o los propios partidos políticos pueden expresar las propuestas de sus programas electorales se han ampliado, dotándose de páginas Web diseñadas expresamente para el periodo de la campaña y perfiles en las redes sociales. Por otro lado, los individuos han visto multiplicadas las fuentes en las que recabar información y expresar su opinión, contando entre otras herramientas de comunicación masiva, con blogs de opciones políticas o de terceros, extensa variedad de periódicos online, y seguimiento de líderes políticos o de opinión (influyentes) en las redes sociales. Es por todo esto que Internet se refrenda una vez más como canal para el análisis de campañas electorales.

Las pasadas elecciones generales de 2010 en el Reino Unido también fueron objeto de estudio en este sentido. El 11 de mayo de ese mismo año, pocos días después de la celebración de estas elecciones, la plataforma de medios Tweetminster (<http://tweetminster.co.uk>) publicaba un documento en Scribd¹³⁸, que bajo el título “Is word-of-mouth correlated to General Election results? The results are in.”, exponía los resultados de su estudio dirigido a identificar si existía una correlación entre el volumen de menciones que recibía los candidatos en Twitter con los resultados electorales¹³⁹. Su metodología es similar a la empleada por Tumasjan et al., (2010), recolectando *tweets* mediante el API de Twitter que mencionan a un

¹³⁸ Scribd es una empresa de publicación y lectura social en Internet. <http://www.scribd.com>

¹³⁹ Se basaron en un análisis llevado a cabo por ingenieros de telecomunicaciones y graduados PHD de la Universidad de Tokyo que trataba de estudiar la correlación entre el “online-buzz” y los resultados electorales de la Elecciones Generales 2009 en Japón. <http://translate.google.com/translate?js=y&prev=t&hl=en&ie=UTF---%208&layout=1&eotf=1&u=http://senkyo.kakaricho.jp/report1.html&sl=ja&tl=en>

candidato y agregando datos (sumando las menciones obtenidas por cada candidato). Durante cuatro semanas de monitorización, capturaron más de 2 millones de mensajes de las 433 circunscripciones (de un total de 650) que obtuvieron representación en esta red social.

Sus conclusiones fueron las siguientes (Tweetminster, 2010, p.2-3)

- El candidato con mayor número de menciones fue el vencedor en cada contienda.
- A mayor volumen de datos, mayor grado de acierto. El modelo se comporta mejor en los ámbitos nacional (2.010.000 tweets) y regional (con una media de 37,000 tweets) que a nivel local (con una media de 667 tweets por circunscripción).
- Tanto el éxito en el acierto como el porcentaje de desviación en la predicción de los resultados electorales es comparable al obtenido por las encuestas.

En esta ocasión, y a diferencia de las advertencias lanzadas posteriormente por Gayo-Avello (2011), las características demográficas y preferencias políticas de los usuarios de Twitter no resultan significativas cuando el volumen de datos es suficientemente grande: *“The accuracy of the predictions in the Twitter experiment were similar to (and better in some cases) the demographically weighted opinion polls. This supports the case that measurements made through data mining in social media channels can be as reliable as traditional opinion polling techniques when the sample size is sufficiently large.”* (Tweetminster, 2010, p.6)

En los últimos años no es difícil encontrar alguna referencia noticiosa que trate sobre el asunto de esta investigación. Así, es posible encontrar dos noticias opuestas publicadas el mismo día, basadas en el mismo “gráfico/análisis” electoral realizado por la CNN¹⁴⁰ junto con la empresa Hexagon con datos extraídos de

140 El gráfico ya no está disponible en

Twitter relacionados con la campaña electoral de 2010 al Senado del estado de Nevada. Los autores del artículo “The 2010 elections: Twitter isn’t a very reliable prediction tool¹⁴¹” informan que aunque el candidato Reid es el favorito en Twitter también ha conseguido un 54% de mensajes negativos en esta red frente al 12% obtenido por su oponente Angle, y que por tanto su victoria se debe a férrea disciplina de partido y estrategia electoral en el logro del voto. De otro lado, en el artículo “Facebook, Twitter Election Results Prove Remarkably Accurate¹⁴²”, se afirma que las encuestas se equivocaron al conceder la victoria a la candidata Angle que perdió con 44,6% de los votos frente al 50,2% de su oponente, predicción en la que Twitter no falló al asignar a Reid un 55% de presencia en la conversación en esta red frente al 45% conseguido por Angle. (Carr, 2010) (Goldstein y Rainey, 2010)

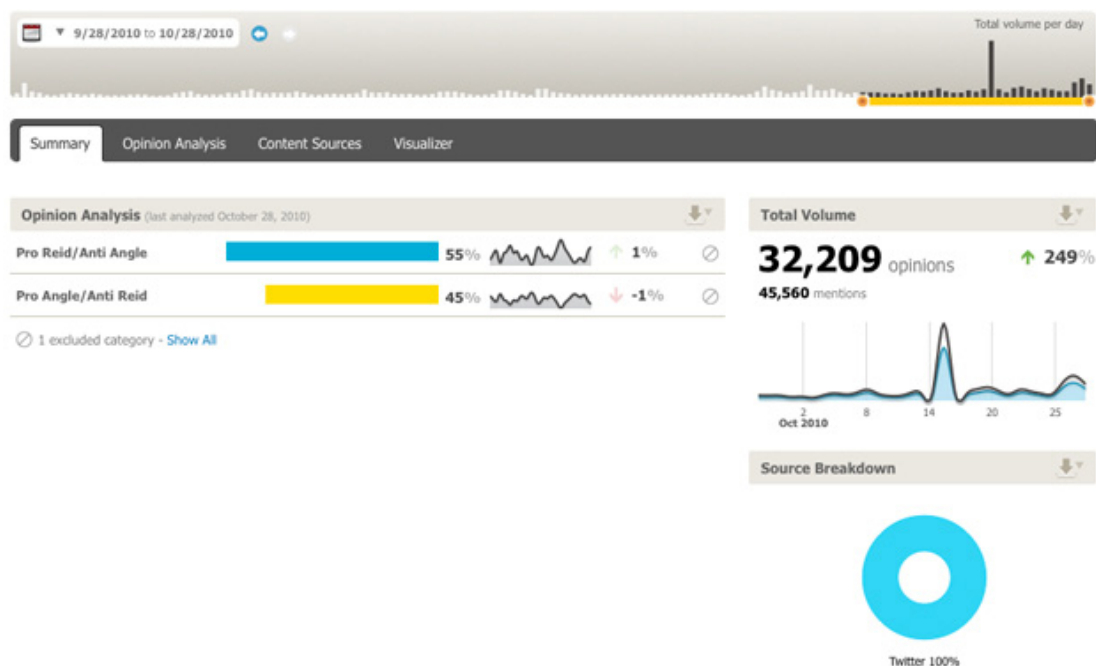
Ilustración 16. The 2010 elections: Twitter isn’t a very reliable prediction tool



<http://edition.cnn.com/interactive/2010/11/election/interactive.nv.senate.pulse/>

¹⁴¹ http://latimesblogs.latimes.com/the_big_picture/2010/11/the-2010-midterms-twitter-effect-not-a-very-reliable-election-prediction-tool.html

¹⁴² <http://www.fastcompany.com/1699853/facebook-twitter-election-results-prove-remarkably-accurate>

Ilustración 17. Facebook, Twitter Election Results Prove Remarkably Accurate

Se ha podido comprobar que existe un cierto grado de contradicción entre los resultados obtenidos en los estudios mencionados, acerca de la capacidad de Twitter para anticiparse a la escena real. Mientras que en algunos de los casos mencionados, Twitter se ha revelado como herramienta de predicción o, cuanto menos, de aproximación a lo que realmente sucede en la “vida real”, mostrando las preferencias políticas de sus usuarios, en otros, por el contrario, su predictibilidad o reflejo de las tendencias sociopolíticas manejan desviaciones considerables y, a veces, alejadas de lo asumible en un estudio con carácter investigador. Bien pudieran estas disparidades ser debidas a la novedad de todo el entramado que significa la introducción de esta Red Social (puesta en marcha, usuarios conectados, análisis de sus datos, etc.). Es por eso que, como ya hemos comentado en ocasiones anteriores, son necesarios estudios que con diferentes objetos y llevados a cabo sobre diferentes entornos, vayan delimitando las variables óptimas a tener en cuenta a la hora de desarrollar una investigación en este entorno.

3.3.4. Análisis de mercados de apuestas como predicción electoral

Varios estudios académicos han confirmado el valor de los mercados de predicción en el pronóstico de próximos eventos, para los resultados electorales, demostrando su eficacia desde las elecciones de finales del XIX y las primeras elecciones del siglo XX. Rhode y Strumpf (2004) investigaron en este sentido 15 elecciones entre 1884 y 1992 en los Estados Unidos, concluyendo que las probabilidades resultantes de las apuestas han demostrado ser notablemente proféticas y casi siempre predijeron correctamente los resultados electorales con suficiente antelación, a pesar de la ausencia de encuestas científicas en algunas de ellas: *“Wagering on political outcomes has a long history in the United States. As Henry David Thoreau noted in 1848, «All voting is a sort of gaming, (...) and betting naturally accompanies it» (Thoreau, 1967, p. 36). This paper analyzes the large and often wellorganized markets for betting on presidential elections that operated between 1868 and 1940. Over \$165 million (in 2002 dollars) was wagered in one election, and betting activity at times dominated transactions in the stock exchanges on Wall Street.”* (Rhode y Strumpf, 2004, p.1)

Con similar metodología de base, examinando 74 batallas electorales en los últimos 130 días de la campaña 2008 en EEUU y recolectando datos de sitios Web de como PollingReport.com, Pollster.com, RealClearPolitics.com e Intrade.com, David Rothschild exploró la precisión de las previsiones derivadas de dos tipos diferentes de datos: las encuestas y los mercados de predicción. Las conclusiones de su estudio confirman los resultados logrados por estudios similares, como el de Rhode y Strumpf (2004).

“In 2008, FiveThirtyEight¹⁴³, a debiased poll-based forecast, offered to the general public a more accurate forecast than raw

¹⁴³ <http://fivethirtyeight.blogs.nytimes.com/>

poll numbers or raw prediction market. But, the analysis here shows that were Intrade's prices debiased, they would have provided a more accurate forecast and more valuable information than the best poll-based forecasts currently available, especially early in the cycle and in uncertain races.”
(Rothschild, 2009, p.913)

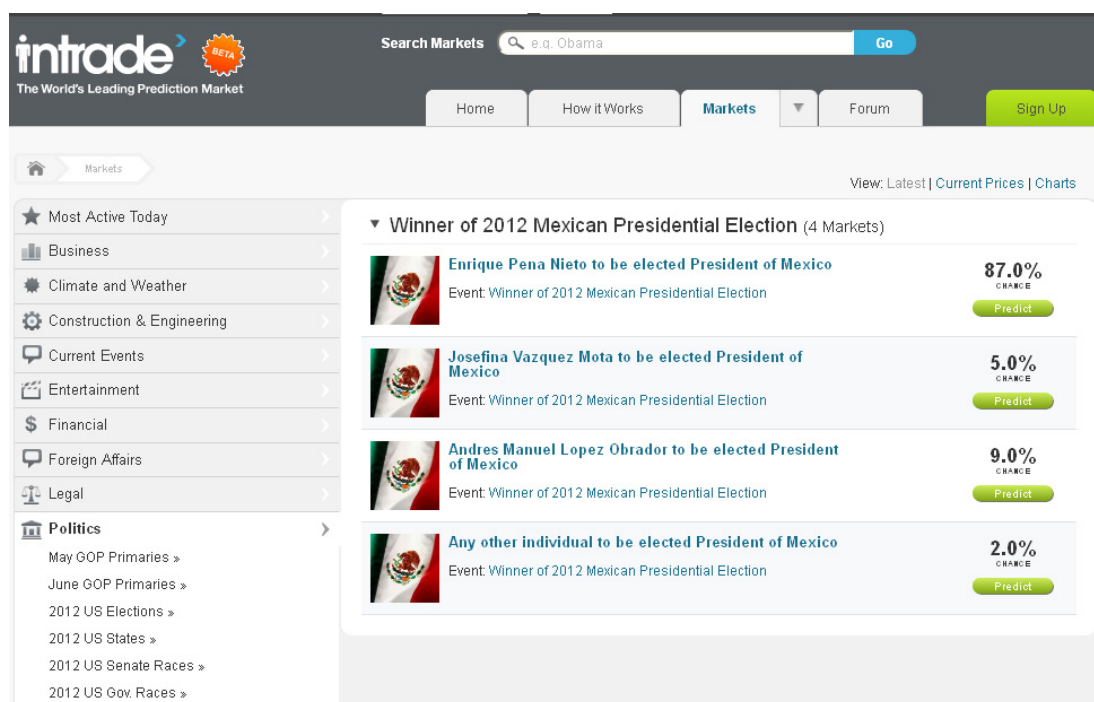
Ilustración 18. Captura de la página Web de Intrade en la que se presenta la predicción electoral en EEUU 2012 en una fecha determinada¹⁴⁴

The screenshot shows the Intrade website interface. At the top, there is a search bar with the text "Search Markets" and a dropdown menu showing "e.g. Obama". Below the search bar are navigation links: "Home", "How it Works", "Markets", "Forum", and a "Sign Up" button. The main content area is titled "Markets" and features a sidebar on the left with categories: "Most Active Today", "Business", "Climate and Weather", "Construction & Engineering", "Current Events", "Entertainment", "Financial", "Foreign Affairs", "Legal", and "Politics". The "Politics" category is selected, showing a list of events: "May GOP Primaries", "June GOP Primaries", "2012 US Elections", "2012 US States", "2012 US Senate Races", and "2012 US Gov. Races". The main content area displays "Most Active in Politics (10 Markets)" with a list of events and their corresponding odds. Each event includes a small profile picture, the event title, the event description, the odds, and a "Predict" button.

Event	Odds
Mitt Romney to be elected President in 2012 Event: 2012 Presidential Election Winner (Individual)	39.2% CHANCE
Sarah Palin to be Republican VP nominee in 2012 Event: 2012 Republican VP Nominee	0.4% CHANCE
Mike Huckabee to be elected President in 2012 Event: 2012 Presidential Election Winner (Individual)	0.2% CHANCE
Bobby Jindal to be Republican VP nominee in 2012 Event: 2012 Republican VP Nominee	4.5% CHANCE
Hillary Clinton to be elected President in 2012 Event: 2012 Presidential Election Winner (Individual)	1.5% CHANCE

144 <http://www.intrade.com/v4/markets/?searchQuery=politic>

Ilustración 19. Captura de la página Web de Intrade en la que se presenta la predicción electoral en México 2012 en una fecha determinada¹⁴⁵

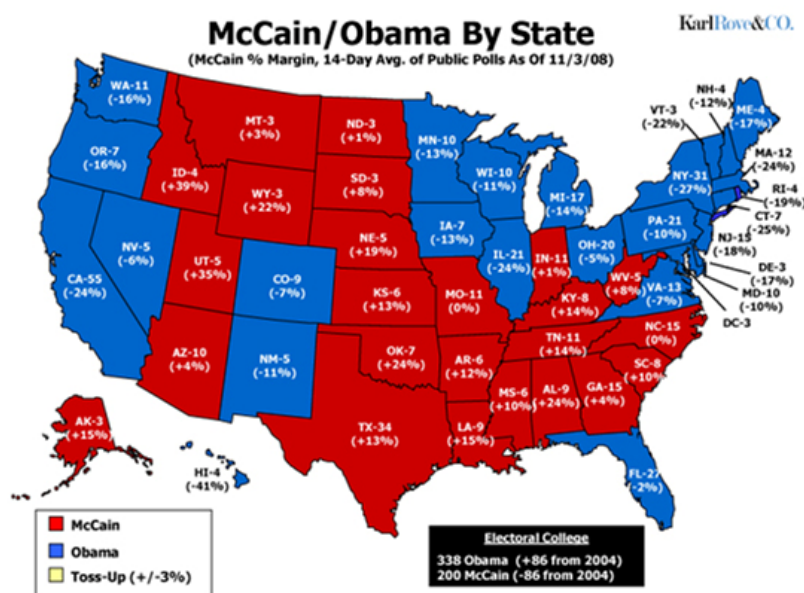


Con similar metodología, Nate Silver¹⁴⁶ un estadístico del béisbol que se inició el análisis de las encuestas políticas, presentó en marzo de 2008 su sitio *FiveThirtyEight.com*, en el que utiliza su propia fórmula para predecir los resultados federales y estatales de las elecciones presidenciales de 2008 en EEUU, prediciendo la victoria de Barack Obama.

¹⁴⁵ <http://www.intrade.com/v4/markets/?searchQuery=politic>

¹⁴⁶ Nate Silver. Estadístico del béisbol que se inició el análisis de las encuestas políticas. En su sitio Web *FiveThirtyEight.com* utilizó su propia fórmula para predecir los resultados federales y estatales y ejecutar las posibilidades de día de las elecciones sobre la base de una serie de factores. Existían entonces otros sitios en los que se combinaban las encuestas, *RealClearPolitics* en particular, pero *FiveThirtyEight* atrajo casi cinco millones de páginas vistas en el día de las elecciones, se convirtió en una de las estrellas.

Ilustración 20. Mapa de la predicción realizada por Nate Silver en las elecciones presidenciales de EEUU en 2008¹⁴⁷



Con motivo de las elecciones EEUU 2012, científicos de datos de Yahoo¹⁴⁸ empezaron a utilizar los mercados de apuestas¹⁴⁹, junto con las encuestas, el análisis de los sentimientos en Twitter, y las tendencias de búsqueda en las consultas para crear una poderosa herramienta predicción de resultado electoral (*The Signal*)¹⁵⁰. El proyecto toma en cuenta soluciones Web basadas en los mercados de predicción, como *Intrade*¹⁵¹ y *OddsChecker*¹⁵², en los que un número considerable de personas

¹⁴⁷ http://www.huffingtonpost.com/2008/11/03/rove-nate-silver-maps-pre_n_140752.html

¹⁴⁸ Yahoo! Inc. es una empresa global de medios con sede en Estados Unidos fundada en enero de 1994 por dos estudiantes de postgrado de la Universidad de Stanford. www.yahoo.com

¹⁴⁹ Un mercado de predicción es un lugar donde los *handicappers* expresan convicciones con dinero real. El precio se determina por la cantidad de usuarios que están dispuestos a invertir dinero con el fin de volver a ganar un dólar si tienen razón. Si una inversión de 90 centavos se requiere para ganar un dólar, los usuarios piensan que el candidato tiene el 90 por ciento de probabilidades de ganar. Si, por el contrario, sólo se requiere una inversión de unos pocos centavos para ganar un dólar, los usuarios están diciendo que las posibilidades del candidato son remotas. Los usuarios apuestan sobre la base de cualquier información de que dispongan, incluyendo datos de próximos eventos, como las encuestas y resultados electorales anteriores, entre otros.

¹⁵⁰ <http://news.yahoo.com/blogs/signal/>

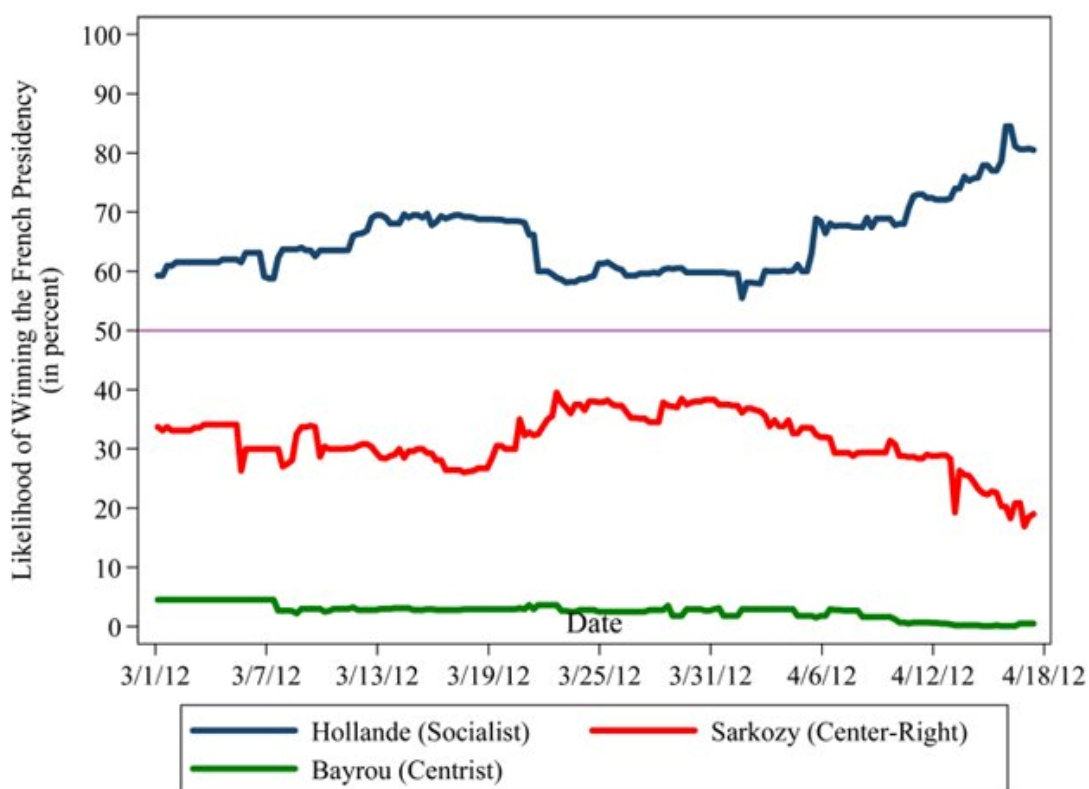
¹⁵¹ <http://www.intrade.com/v4/markets/?eventClassId=19>

¹⁵² <http://www.oddschecker.com/specials/politics-and-election>

apuestan sobre los resultados de elecciones próximas a suceder o que se encuentran en periodo de campaña. (Rothschild, 2012)

Concretamente, con fecha del 19 de abril de 2012 publicaron sus estimaciones de victoria para las elecciones francesas basándose en datos de *Intrade* y *OddsChecker*, resultado de las cuáles el candidato socialista François Hollande aparecía como favorito frente a la reelección del entonces presidente Nicolas Sarkozy.

Ilustración 21. Estimación de la victoria de las elecciones francesas 2012, utilizando mercados de predicción/apuestas, de Rothschild (2012)



Likelihood of Winning French Presidency

Sources: Intrade and OddsChecker

El 24 de abril el diario español El País publicaba los resultados de la primera vuelta, (la primera vuelta se celebró el 22 de abril de 2012, la segunda, el 6 de mayo

de 2012): 28,63% para François Hollande y un 27,18% para Nicolas Sarkozy¹⁵³. El resultado es acorde con el obtenido con las técnicas de seguimiento de los mercados de predicción difundido por Yahoo.

El 24 de octubre de 2012, el artículo “Oct. 23: The Virtues and Vices of Election Prediction Markets”¹⁵⁴, publicado en el espacio de *FiveThirtyEight* en el diario estadounidense *The New York Times*, Nate Silver, a tenor de las 11 encuestas publicadas sobre las posiciones de los dos candidatos a la Presidencia de los Estados Unidos de América (Barack Obama y Mitt Romney), las cuales, en conjunto presentan un escenario neutral (5 a favor de Obama, 4 a favor de Romney y 2 empate), realiza un estudio comparativo entre sus propias predicciones y las conseguidas por las empresas de mercados de apuestas como Intrade, OddsChecker y Betfair (y *sport books* o casas de apuestas deportivas). Mientras que en general, los datos de éstas y los suyos se encuentran más cercanos entre sí que los arrojados por las encuestas tradicionales, existe una diferencia considerable marcada por Intrade (para victoria de Obama, 68%- *FiveThirtyEight* y 63%-Betfair, versus 55%-Intrade). El autor argumenta las siguientes razones para la existencia de dichas desviaciones:

- Las estimaciones de *FiveThirtyEight* se basan exclusivamente en análisis de información pública, mientras que los mercados de apuestas pueden utilizar ésta junto con las predicciones de sus competidores para calibrar sus resultados.
- Las apuestas de Intrade son de precio bastante más elevado que en el resto. El hecho de que estos márgenes relativamente amplios existan entre los diferentes mercados sugiere que son relativamente *illiquid*¹⁵⁵ y que conllevan altos costos de transacción, lo que puede limitar su eficacia.

153 http://internacional.elpais.com/internacional/2012/04/17/actualidad/1334680532_668710.html. (Consultado el 02/05/2012)

154 <http://fivethirtyeight.blogs.nytimes.com/2012/10/24/oct-23-the-virtues-and-vices-of-election-prediction-markets/>

155 El estado de un activo de seguridad o de otro tipo que no puede ser vendido o canjeado por dinero en efectivo sin una pérdida sustancial de su valor. Los activos no líquidos tampoco se puede vender

- La diferencia en la fiabilidad o sofisticación de los “apostadores”. Los de Intrade no son necesariamente los más preparados.
- Intrade se cita con más frecuencia por los medios de comunicación estadounidenses. Este hecho abre la posibilidad de que alguien pueda hacer una apuesta sobre uno de los candidatos con el fin de influir en las percepciones de los medios acerca de qué candidato tiene el *momentum*.

Los ejemplos incluidos en este epígrafe asientan la idea de que, hoy en día, gracias a Internet y las nuevas tecnologías, son posibles métodos de predicción alternativos a las encuestas preelectorales tradicionales. Dicha afirmación viene a consolidar tanto el objeto como el método de nuestra investigación: el análisis de la red Twitter puede ser capaz de predecir resultados electorales.

3.3.5. El peso de las Ciber campañas en el éxito electoral

Si bien la aparición de Howard Dean en 2003 ha sido señalada como *the coming of age* en el mundo de las campañas en Internet, fueron las elecciones de 1996 en EEUU la primera auténtica ciber campaña desarrollada en el marco de una campaña electoral, en la que ambos candidatos, Bob Dole y Bill Clinton manejaron sus Webs de campaña¹⁵⁶. En Europa, son las elecciones generales inglesas de 1997 las que han sido designadas como la “primera elección en Internet”, ya que en ellas un gran número de partidos políticos utilizaron las nuevas tecnologías con el fin de atraer a los votantes y hacerlos partícipes de su política¹⁵⁷.

rápidamente debido a la falta de inversores dispuestos y dispuestas o especuladores para comprar el activo. La falta de compradores dispuestos también conduce a las grandes discrepancias entre el precio de venta (del vendedor) y el precio de licitación (de un comprador).

¹⁵⁶ MARGOLIS, Mike, RESNICK, David y WOLFE, Joel (1999). “Party Competition on the internet: Minor Versus Major Parties in the UK and USA”, *Harvard International Journal of Press/Politics*, 4 (4): 24-47.

¹⁵⁷ WARD, Stephen J. y GIBSON, Rachel K. (1998). “The First internet Election? UK Political Parties and Campaigning in Cyberspace”, *Ivor Crewe, Brian Gosschalk and John Bartles (eds.) Political Communications: The General Election Campaign of 1997*.

A lo largo de estas décadas, varios autores han estudiado el impacto de la página web de un candidato en su resultado electoral. D'Alessio¹⁵⁸ en su análisis de las elecciones al Congreso de EEUU en 1996 ya concluía que contar con una cibercampaña proveía al candidato con 9,3 más votos de los previstos inicialmente, valores únicamente superados por la variable “*incumbency*”, y por la posición más o menos relevante que éste ocupa dentro de su partido. D'Alessio fue escéptico con sus propios resultados, no confiando plenamente en que sólo por el hecho de disponer de una página Web, una opción política obtuviera sin más, un considerable número adicional de votantes. Profundizando en las causas de estos buenos resultados, el autor argumenta que la variable Web es consecuencia del grado de preparación y profesionalismo de la campaña (Gibson y McCallister, 2005, p.6):

- Cuanto más sofisticados son los recursos de un candidato, más probabilidad existe de que disponga de página Web.
- La página Web no hace que los electores cambien de opinión sino que moviliza a los votantes que sin su presencia se hubieran abstenido.
- Los candidatos que ocupan las primeras posiciones en las encuestas fijarían su atención en el beneficio añadido de poner en marcha su página Web.

Diversos autores¹⁵⁹ han estudiado posteriormente el efecto de las cibercampañas en los electores llegando, en la mayoría de los casos, a la conclusión opuesta a la de D'Alessio. En general, las páginas Webs de los candidatos u opciones políticas sirven para reforzar las ideologías de aquellos que las consumen y quizás movilizar a una pequeña cantidad de votantes indecisos, pero que en raras ocasiones

¹⁵⁸ D'ALESSIO, David, W. (1997). “Use of the Web in the 1996 US Election”. *Electoral Studies* 16(4): 489-501. Mencionado por Gibson y McCallister (2005).

¹⁵⁹ CARLSON, T. (2006) “It’s a Man’s World”? Male and Female Election Campaigning on the Internet,” *Journal of Political Marketing*, 6, 41-67

FARNSWORTH, S. J. y OWEN, D. (2004) “Internet use and the 2000 presidential election”, *Electoral Studies* 23(3): 415-429.

BIMBER, B. y DAVIS, R. (2003) “Campaigning Online: The Internet and US Elections”. *Oxford University Press: Oxford*.

procuran un cambio de preferencia de voto. Pippa Norris (2003) aseveró esta afirmación al recoger en una de las conclusiones de su estudio que los recursos empleados en las páginas Web's de los partidos políticos en Europa servían antes que para ampliar la participación de grupos con bajo interés o con cierto grado de desafección hacia la política, para alcanzar a los ciudadanos políticamente activos, interesados y comprometidos: *"Like traditional news media, politics on the Internet serves primarily to reinforce civic engagement (Norris, 2000)."* (Norris, 2003, p.43)

En este momento conviene apuntar que uno de los *main targets* de las campañas electorales consiste en fidelizar a los votantes predispuestos y movilizarlos evitando la abstención en la jornada electoral, junto con la necesidad de atraer y convencer al mayor número de indecisos, articular la programación de una hábil cibercampaña capaz de aproximar a cualquier individuo a nuestra causa a través de las múltiples herramientas de que se dispone en Internet (consultas/visitas a la página Web, registro de usuarios para envío de emails, fan en Facebook, seguidores en Twitter, etc.), supone multiplicar o al menos, intensificar la probabilidad de éxito. Por tanto, sería posible deducir que el menor de los contactos, establecido a través de cualquier medio al alcance, con un votante incrementaría porcentualmente la probabilidad de éxito electoral.

El mismo autor, cuya investigación perseguía teorizar si dichas páginas cumplían la función de comportarse a modo de foro cívico plural y canales para la participación política, pudo afirmar que, además de fortalecer el pluralismo, las páginas Web's posibilitaban a los partidos minoritarios dotarse de una visibilidad incapaz de conseguir en los medios de comunicación: *"The availability of party websites has strengthened communication pluralism in Europe by widening the information available about minor and fringe parties, allowing them greater voice and visibility than coverage in traditional news media."* (Norris, 2003, p.42)

En este punto, nos encontramos con opiniones contrarias, unas favorables al efecto de las cibercampañas en los electores y otras que les restan valor e incluso les atribuyen un valor nulo. Estas contradicciones observadas, movieron a Gibson y

McCallister a proceder con un nuevo estudio con el ánimo de conocer el verdadero efecto de una campaña en la Red: *Overall, therefore, despite the fact that the audience for web campaigning is undergoing significant expansion, our understanding of whether it actually produces any tangible results in terms of increasing levels of voter support is far from clear.*” (Gibson y McCallister, 2005, p.8)

Gibson y McAllister estudiaron el efecto de las cibercampañas en el resultado electoral en las elecciones australianas de 2004. Los investigadores concluyeron que una cibercamapaña es un elemento primordial a la hora de asegurar la victoria electoral. El uso de una página Web proporciona un 4% adicional a la preferencia de voto inicial, lo cual supone más que todas las variables que concurren en la campaña, excepto la situación de re-elección (*incumbency*): *...we examine whether candidates' web use or cyber-campaigning affects their fortunes at the ballot box, controlling for a range of personal, partisan and other campaign related activities. The findings show that having a campaign website does have a significant and positive independent effect on a candidate's levels of support.*” (Gibson y McCallister, 2005, p.3)

Para llevar a cabo su investigación, analizaron los siguientes aspectos (Gibson y McCallister, 2005, pp.9-15):

- La tendencia de los candidatos a poner en marcha su página Web en campaña electoral, evaluada en función de las variables recursos o atributos personales (edad, género, educación superior), partido político (Labor, Democrat, Green, One Nation) y posición política (re-elección - *incumbency*-, experiencia legislativa, antigüedad en el partido). Los atributos personales no influyen, la experiencia (como es el caso de ser un candidato que se presenta a la re-elección) es un hecho significativamente influyente, y contrariamente a lo que pudiera esperarse, no son los grandes partidos los que lideran el ranking de

páginas Web en campaña, sino que el partido verde emerge como el más dispuesto a la comunicación online.

- El uso de internet por votantes y candidatos: sólo un 18% de los votantes ha accedido a Internet para informarse sobre la campaña. Además, comparando candidatos y votantes, aquellos se mostraron más propensos a consumir información relacionada sobre las elecciones.
- La conexión existente entre disponer de página Web y el resto de actividades de campaña¹⁶⁰: la producción de una cibercampaña parece más un lujo que una necesidad, siendo aquellos candidatos que han preparado su campaña con antelación los más dispuestos a dotarse de página Web, así como aquellos que disponen de más tiempo y de personal colaborador.
- El incremento del número de votos sobre las primeras estimaciones, en aquellos candidatos que cuentan con página Web, comparando todas las variables más la variable “Página Web” con el voto: el uso de una página Web proporciona un 4% adicional a la preferencia de voto inicial.
- Conocer si los votantes que acceden a Internet para informarse acerca de la campaña presentan comportamientos o actitudes de campaña diferentes al resto de votantes, estableciendo tres grupos diferenciados¹⁶¹: el grupo que siguió las elecciones en Internet mostró una mayor independencia ideológica, siendo más probable que consideraran cambiar su voto hacia un candidato diferente del previsto inicialmente.

Tabla 10. Relación entre condiciones particulares y recursos empleados por candidatos en sus páginas Web's y su peso en relación al resultado electoral en las

¹⁶⁰ Actividades de campaña (entrevistas en los medios, actividades dentro del partido y recaudación de votos), apoyo del partido (soporte organizativo y el número de trabajadores de que se dispone durante la campaña), y fecha de inicio de preparación de la campaña.

¹⁶¹ Votantes sin acceso a Internet, votantes con acceso a Internet pero que no siguieron las elecciones en este medio, votantes que siguieron las elecciones en Internet.

elecciones australianas de 2004, extraída del trabajo de Gibson y McCallister (2005)

Table 5: Resources, the Campaign, Candidates' Web Page, and the Vote

	Means	OLS coefficients	
		b	beta
<i>Web page</i>	.36	4.05**	.11**
<i>Individual resources</i>			
Age (years)	57.64	-.01	-.02
Gender (male)	.67	.19	.01
Tertiary education	.89	2.00	.04
<i>Party membership</i>			
Labor	.20	-6.12**	-.13**
Democrat	.22	-28.35**	-.64**
Green	.28	-25.16**	-.62**
One Nation	.12	-26.85**	-.48**
<i>Political resources</i>			
Incumbent	.13	12.98**	.24**
Legislative experience	.11	1.32	.02
Length party membership (years)	9.36	-.15**	-.08**
<i>Campaign activities (hours per week)</i>			
Media interviews	3.72	-.15	-.04
Canvassing	11.16	.04	.03
Party activities	15.85	.08*	.07*
<i>Party support</i>			
Organizational support (0-10)	5.50	.10	.01
Party workers (number)	6.55	.00	.00
<i>Length of campaign preparation (months)</i>	7.51	.34*	.08**
Constant		26.08	
Adj R-sq		.67	
(N)		(420)	

** statistically significant at $p < .01$, * $p < .05$, two-tailed.

OLS regression estimates showing partial (b) and standardized (beta) coefficients predicting the percentage first preference vote for House of Representatives candidates only ($n = 420$). All variables are scored zero or one except age (years) and length of party membership (years), campaign activities (hours per week) organization support (0-10), party workers (number) and length of campaign preparation (months). For party membership, the excluded category is Liberal-National candidates.

Source 2004 ACS.

Los resultados de estas investigaciones vienen a reforzar la idea de que contar con una página Web es un componente importante a la hora de tratar de asegurar un mayor éxito electoral, dado que refuerza a las bases, moviliza a los abstencionistas y puede cambiar la tendencia de voto en electores ideológicamente independientes.

Bien es cierto que las investigaciones incluidas en este epígrafe analizan procesos electorales sucedieron bastante tiempo atrás y que actualmente, son extraños los casos en los que las principales opciones políticas que concurren en un proceso electoral (al menos en el mundo occidental), carezcan de una campaña online más o menos desarrollada, acorde con los medios o presupuestos disponibles y con la confianza propia depositada en este canal. Esta afluencia masiva, constataría tanto los resultados de las investigaciones precedentes como la tendencia de la población a informarse en Internet y a participar en las cibercampañas.

3.4. Encuestas preelectorales. Eficacia y desviación

“Tras tantas elecciones seguidas estimando acertadamente el resultado más probable, este claro tropiezo bastó para poner nuevamente en cuestión los sondeos electorales como instrumentos fiables para la detección de estados de opinión.”(Toharia, 2012)

A medida que las campañas electorales se fueron profesionalizando, desarrollando los candidatos estrategias cada vez más elaboradas, dedicando mayor cantidad de recursos de toda índole a su desarrollo y apoyándose en la capacidad de difusión de los entonces nuevos medios de comunicación de masas, la evolución de la medición de la opinión pública mediante sondeos ha mostrado un signo paralelo de relevancia en el entorno de las campañas.

La importancia de las encuestas en la política es fácil de deducir simplemente atendiendo las noticias en un día determinado. Esto es aún más cierto en periodos de elecciones de carácter nacional, cuando los medios de comunicación inundan sus portadas diarias con los resultados de las últimas encuestas, las cuales se nos presentan como una instantánea real servida como predicción de los resultados electorales a los que hacen referencia.

Manteniendo su origen en los estudios de mercado, dedicados al conocimiento de las tendencias o preferencias que los consumidores mostraban hacia tal o cual marca comercial y dirigidos a obtener valores que indicaran a las empresas y productores de bienes y servicios tanto a tratar de averiguar lo que los consumidores preferían e incluso esperaban o les gustaría, como sus consideraciones y críticas directas sobre un producto o servicio ya dispensado en el mercado, los estudios de mercado preelectorales, surgieron y se desarrollaron con el advenimiento de campañas electorales sucesivas que presentaban con el tiempo un nivel de profesionalidad *in crescendo*. Como parte de esta profesionalización de la planificación y estrategias de campaña, los estudios preelectorales (bien en campaña o antes de ella) y los estudios postelectorales, se han convertido en una pieza del mayor interés (y de relevancia, ya sean de carácter privado dedicado a los partidos, o público, con difusión general-social) en el marco de la planificación de una estrategia de campaña (del lado de partidos políticos y candidatos, generalmente lideradas por consultoras más o menos afines al partido que se trate) así como de divulgación democrática de una campaña electoral (del lado de la retroalimentación de los valores de la democracia en las sociedades libres, generalmente lideradas por institutos de análisis de la opinión pública)¹⁶².

Autores reconocidos como Dennis Kavanagh, fijan las elecciones de 1936 en EEUU, en las que los candidatos Franklin Delano Roosevelt como candidato del Partido Demócrata y Alf Landon como candidato del Partido Republicano, disputaban la elección a la presidencia. En esta campaña se realizó la primera encuesta electoral masiva de George Gallup¹⁶³, que con una muestra de 5,000 individuos interrogados por todo el país predijo en que el vencedor de las elecciones

¹⁶² Las encuestas que cuentan con mayor difusión pública en nuestro país son las realizadas por el CIS (Centro de Investigaciones Sociológicas) y las desempeñadas por institutos o centros de estudios privados a comanda de diversos medios de comunicación (Metroscopia, Ipsos, etc.).

¹⁶³ Periodista, matemático y estadístico estadounidense. Fundó el *American Institute of Public Opinion*, Instituto de Opinión Pública estadounidense, en 1935 para realizar sondeos electorales en Estados Unidos, conocer los gustos de la gente y estudiar la opinión de la masa social. Fue el pionero en la medición de la audiencia tanto de radio como de televisión mediante encuestas.

sería el demócrata Roosevelt. En paralelo la revista *The Literary Digest*, cuyo error ya fue mencionado en el epígrafe dedicado a Twitter, concretamente en el biblioreferenciado estudio de Gayo-Avello (2011), realizó una encuesta pidiendo a los suscriptores de su revista enviar cupones anónimos indicando su candidato preferido en las elecciones, recibiendo dos millones de respuestas y dando como vencedor a Alf Landon¹⁶⁴. Esta fatal predicción de *The Literary Digest*, con unos resultados realmente alejados de la realidad final, fue el origen de su decadencia y posterior desaparición de escena.

Con todo, las encuestas de opinión, basadas en cuestionarios aplicados a una muestra representativa del electorado, configurada sistemáticamente, entraron a tallar en la escena política ya a contar de la década de los treinta. Una fecha clave es la de 1936, cuando la encuesta Gallup y las encuestas de Elmo Roper predijeron de manera acertada los resultados de la contienda presidencial Roosevelt-Landon en los Estados Unidos.
(Kavanagh, 1994, p.2)

Dejando a un lado, dado que no es objeto de esta investigación, el interés o persecución en dilucidar si la difusión pública de los resultados de las encuestas preelectorales (máxime si éstas se llevan a cabo en periodo de campaña electoral) dispone de la capacidad de influir en el comportamiento final del electorado¹⁶⁵, esto es, en su decisión última de voto, el que una encuesta “dé en el clavo”, que prediga con ventaja temporal que un candidato resultará o no vencedor en una elección dada, ha sido objeto de estudio que se ha incrementado en los últimos años con los

¹⁶⁴ Elecciones presidenciales de Estados Unidos de 1936. (http://es.wikipedia.org/wiki/Elecciones_presidenciales_de_Estados_Unidos_de_1936. (Consultado el 10/12/2011))

¹⁶⁵ El impacto de la difusión pública de las encuestas podría afectar a la decisión final del votante. No obstante, Kavanagh concluye que “no hay hallazgos concluyentes respecto al impacto de las encuestas en los votantes”. (Kavanagh 1994:2) No obstante, escribe que se sugieren varios efectos posibles: las encuestas pueden impulsar a los partidos minoritarios, pueden alentar el voto táctico y pueden indicar a los votantes algo acerca de la efectividad de su voto. (Kavanagh 1994:158)

primeros avances de estudios sobre su precisión y el sentimiento popular que se ha desprendido, una vez conocidos los resultados electorales, sobre la tendencia a la baja en la fiabilidad de los pronósticos de las encuestas. La confianza en la fiabilidad y el acierto de las encuestas (más allá del margen de error que las mismas dan por supuesto y el nivel de confianza que se calcula, en relación a los resultados electorales definitivos) ha devenido, de un tiempo a esta parte, en un “estado de la cuestión”, que se ha vuelto un tanto “cuestionable”¹⁶⁶.

Ya en el estudio de Kavanagh (1994) se enumera una amplia muestra de variaciones en el éxito predictivo de las encuestas como herramienta de predicción del resultado electoral: *“Justa o injustamente, las encuestas de opinión son generalmente evaluadas por la precisión de sus proyecciones.”* (Kavanagh, 1994, p.7) Y como él mismo comenta, la tarea del pronóstico no es una labor fácil de llevar a cabo con éxito, *“La predicción es una labor en extremo riesgosa para los encuestadores. A la luz de las dificultades que ella plantea, es sorprendente que operen tan bien como lo hacen.”* (Kavanagh, 1994:152)

Lau (1994), analizando las encuestas de las elecciones presidenciales estadounidenses de 1992 en busca de los posibles factores que inciden en su mayor o menor grado de precisión, encontró que factores como el número de días en que se llevó a cabo la encuesta, mejoraban la exactitud de las encuestas. De la misma manera, halló que algunos de los factores que se suponen asociados con una mayor precisión de la encuesta, como el tamaño de la muestra, el muestreo con filtros (los votantes probables frente a los no probables), porcentaje de indecisos, y la cercanía o lejanía en el que se llevó a cabo la encuesta en relación con el día de la jornada electoral, no resultan significativos.

¹⁶⁶ El mayor valor predictivo de las mismas reside en su capacidad para realizar “proyecciones de voto”, esto es, concluir a raíz de los datos obtenidos en su estudio, cómo se comportará el electorado y a qué opción votarán.

Casi diez años después, Desart y Holbrook (2003), tomando como partida las conclusiones del estudio de Lau (1994), profundizaron en su análisis comparado de las encuestas realizadas en las elecciones estatales de 1992, 1996 y 2000 en Estados Unidos¹⁶⁷:

“Using Lau’s (1994) examination of national-level polls as a guide, we find that sample-related characteristics like sample size and the use of a likely voter sampling filter tend to improve the predictive accuracy of these polls. More importantly, however, we find that the number of days to election has an important impact on poll accuracy and that this effect can vary substantially across different campaign contexts.”

Los posibles factores de error que tuvieron en cuenta y en los que basaron su investigación comparada, son los siguientes:

- Tamaño de la muestra. Por definición, se espera que a mayor tamaño de la misma el margen de error se reduzca.
- Votantes probables. Aquellos que afirman van a votar y que por tanto dibujarían un universo más reflexivo (y efectivo).
- Días antes de la elección. Asumiendo que una de las consecuencias de las campañas electorales es cambiar las opiniones de los votantes, se estima que las encuestas realizadas en una fecha cercana a la jornada electoral reducen su margen de error.
- Número de votantes indecisos. Se da por hecho que a un número mayor de encuestados definidos como votantes indecisos, le corresponde un menor grado de precisión en los resultados de las encuestas

¹⁶⁷ De un total de 244 de casos en 1992, 207 en 1996 y 380 en el año 2000, se eliminaron aquellos en los que no se disponía de información completa, quedando 232 casos en 1992 y 145 en 1996.

- Jornadas de campo. Siguiendo a Lau (1994), se espera que a mayor duración en la captura de las respuestas en una encuesta dada (o toma de la muestra) las encuestas presentarán un margen de error mayor.
- Fin de semana. Lau (1994) presumía que los contactos encuestados desarrollados en el fin de semana (los encuestados estaban en su casa o fuera del trabajo) obtendrían mejores precisiones que las encuestas realizadas a lo largo de la semana.
- Contexto electoral. Asumen que las circunstancias específicas que acontecen a lo largo de una campaña electoral pueden incidir en una mayor o menor precisión de las encuestas.

Sus conclusiones directas se concretan en la siguiente tabla:

Tabla 11. Factores influyentes en el margen de error de las encuestas

Influyentes	Poco o nada influyentes
<ul style="list-style-type: none"> ▪ Tamaño de la muestra ▪ Votantes probables ▪ Contexto electoral ▪ Días antes de la elección 	<ul style="list-style-type: none"> ▪ Votantes indecisos ▪ Jornadas de campo ▪ Fin de semana

Fuente: elaboración propia a partir de datos del estudio de Desart y Holbrook (2003)

Refiriéndonos a España, encontramos estudios recientes dedicados a profundizar en el grado de eficacia de las encuestas. El primero que mencionamos es el del investigador Pedro Magalhães, *PhD* en Ciencias Políticas por la *Ohio State University*, que ha dedicado parte de su investigación al seguimiento de la eficacia de las encuestas¹⁶⁸. En una presentación ofrecida a los alumnos del Diploma “*Leadership, Communication and U.S. Politics*” en la *Georgetown University* en

¹⁶⁸ Página personal en <http://sites.google.com/site/pmdccm/>. Cuenta con un blog en el que difunde sus análisis, titulado “Margens de erro” (<http://margensdeerro.blogspot.com.es/>)

octubre de 2011, concluyó que las encuestas españolas no son reconocidas por su brillante reputación, siendo sus mayores errores de predicción los siguientes:

- La media absoluta de desviación entre los resultados estimados y los resultados obtenidos por los partidos mayoritarios (error 3).
- La desviación entre las estimaciones y los resultados para el partido ganador (error 5)

Ilustración 22. Desviaciones encuestas electorales españolas en 2008 según análisis de Pedro Magalhães (2011)

9th March 2008

Pollster	Last fieldwork day	n	PSOE	PP	IU	CiU	ERC	Error 3	
Metroscopia	Feb 27 th	600	42.9	38.8	4.8	2.6	1.9	0.8	0.1
DYM	Feb 25 th	1,074	42.0	40.0	4.1	-	-	0.8	-2.0
Noxa	March 3 rd	-	43.5	39.5	4.4	-	-	0.5	0
Opina	March 2 nd	1,000	43.5	38	5.0	2.7	1.5	0.8	1.5
Demometrica	Feb 26 th	8,000	44.2	38.6	-	-	-	0.8	1.6
Sigma Dos	March 2 nd	1,000	43.4	39.3	4.9	2.9	1.6	0.5	0.1
Obradoiro	Feb 28 th	1,810	45	38.8	4.5	3.1	1.5	0.7	2.2
Average			43.5	39.0	3.8	3.0	1.6		
Election outcome			43.9	39.9	3.8	3.0	1.2		
Average error 3			0.7						
Average error 5			+0.5 (overestimation of PSOE's lead)						

Otras conclusiones resultantes de su análisis de encuestas en campaña electoral en España desde 1989 hasta 2008, relacionadas con las desviaciones entre los resultados estimados por los sondeos y los resultados finales obtenidos, son:

- Las predicciones difieren según se trate de un tipo u otro de encuesta (personal, telefónica, postal).

- Cuando se da cambio de opción política en el Gobierno, el PP (Partido Popular) aparece sobreestimado en las predicciones.
- El diario El Mundo presenta unos resultados mejores para el PP que el diario El País, y éste, mejores proyecciones para el PSOE (Partido Socialista Español) que las noticiadas por El Mundo.

Y realmente de interés, para uno de los casos que se analiza en nuestro muestreo de trabajo (Elecciones Generales 2011 en España), las siguientes predicciones arrojadas con un mes de antelación:

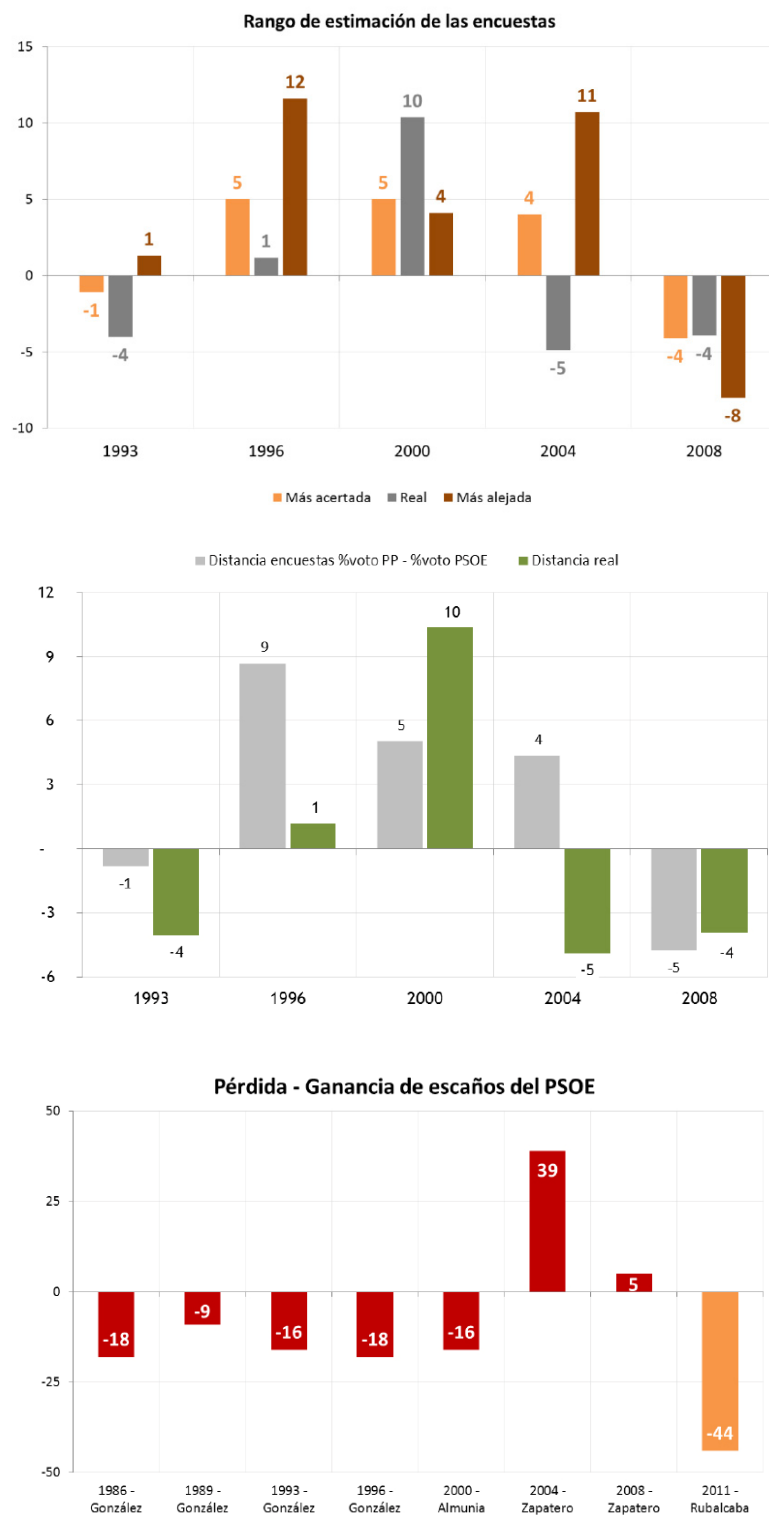
- Resultados ligeramente por encima de los estimados por las encuestas.
- Tercer peor resultado para el PSOE desde su concurrencia a las urnas.
- Peor resultado para un “*incumbent*” desde las elecciones de 1982.

En un segundo estudio, llevado a cabo por la consultora GAD3 y referido a las desviaciones de las encuestas publicadas con motivo de las Elecciones Generales 2011 en España, los autores de preguntan: “¿Cuánto aciertan las encuestas preelectorales en España?”. De su análisis, que tomaba como referencia el conjunto de más de doscientos sondeos desarrollados en nuestro país desde 1993 en comparación con los resultados electorales definitivos, se desprenden las siguientes lecciones: (GAD3, 2011, pp.4-6)

- “La coincidencia de las encuestas no garantiza su precisión”... “la media de las encuestas siempre es menos precisa que la estimación realizada por uno de sus extremos”.¹⁶⁹
- “El voto en el partido en el gobierno suele infraestimarse”... “En todos los casos salvo en 2004, el partido en el gobierno obtuvo mejores resultados de los estimados”.
- “La desafección al PSOE nunca ha sido tan abrupta”.

¹⁶⁹ Auguran que la diferencia entre los partidos mayoritarios PP y PSOE sería o bien menor del 11% o bien superior al 16%, no encontrándose en ese rango. A día de hoy, en que es posible conocer esa diferencia real, los datos son: 44,62% de votos para el PP y 28,73% para el PSOE. http://www.generales2011.mir.es/99CG/DCG99999TO_L1.htm. (Consultado el 07/01/2012)

Ilustración 23. Gráficos de resultados del análisis sobre desviación de encuestas españolas realizado por GAD3



En las recientes Elecciones Andaluzas 2012 en España los errores en las predicciones resultantes de las encuestas preelectorales han sido manifiestamente notables: *“el 25-M pasará a la historia política andaluza como el día en que todo el mundo se equivocó. Los encuestadores, los periodistas, los sociólogos, los políticos, los analistas... ¡Santo Dios, no acertó absolutamente nadie!”* (Avellanado, 2012)

En dichos comicios todas las encuestas concedieron la mayoría absoluta al PP, después de 30 años de gobierno socialista. En ellas, la diferencia en porcentaje de votos entre ambos partidos oscilaba entre un 5% (CIS) y un 11% (El Mundo), y en los resultados electorales la diferencia real fue de tan sólo de 1,14% a favor del primero. La pregunta generalizada tras los comicios fue: ¿en qué fallaron? Los expertos abdujeron que lo más probable es que se debiera a que el voto oculto al partido socialista no afloró en las encuestas y que muchos de los entrevistados sintieran vergüenza a la hora de expresar su intención de votar al PSOE: “Esteban López-Escobar, catedrático de Opinión Pública de la Universidad de Navarra, cree que «pudo darse el fenómeno de la «vergüenza» de votar socialista», tras los escándalos de corrupción en los ERE y la extendida difusión de que el PP ganaría de largo las elecciones.” (Arrizabalaga, 2012)

Dicha afirmación es corroborada por Toharia (2012, pp.62-64) cuando, refiriéndose a: “¿Qué ha pasado en las Elecciones Autonómicas Andaluzas?”, incluye los siguientes motivos:

- La producción de una “espiral del silencio” que propició la aparición de un voto silente y escondido hacia el PSOE.
- Relacionada con la anterior, la producción de un leve pero relevante vuelco de opinión (frente al temor de la anunciada “marea azul”) en los últimos días de la campaña que llevó a los indecisos a inclinarse por este partido.
- Puede que la tan anunciada clara victoria del PP propiciara un exceso de confianza en su electorado, que por tanto, dejó de acudir masivamente a las urnas.

En entornos de sistemas democráticos occidentales como en España, los medios de comunicación (prensa escrita y radio, principalmente) son los mayores clientes en la demanda de realización de encuestas¹⁷⁰. Dado que éstos, prácticamente en su totalidad, se encuentran alineados con alguna de las opciones políticas en pugna, las empresas encuestadoras operan como si de un *outsourcing* se tratara, pudiendo no disponer del control absoluto sobre la cobertura colorida y, en cierto modo, equívoca, que los medios dan a las predicciones obtenidas mediante los sondeos. Esta realidad junto con los errores o desviaciones producidos por las proyecciones de voto manejadas en la trastienda o “cocina” de las encuestas (incluida la asignación de los votantes indecisos y los cambios de la intención de voto), la cautela o reticencia de algunos encuestados a revelar sus tendencias reales de voto¹⁷¹, y la mayor o menor idoneidad y representatividad de la muestra elegida¹⁷², parecen ser las principales causas de su posible fallo en la predicción.

El cálculo de una “muestra representativa”, es decir la selección de la muestra de individuos a los que se va a encuestar, tiene una gran importancia ya que si los entrevistados nos representan al colectivo de electores objeto de la investigación, los resultados serán erróneos¹⁷³. Si la muestra está mal elegida (muestra sesgada), se dice que no es representativa, y puede producir errores imprevistos e incontrolados, conocidos como “errores de muestreo”. Por lo general, en las encuestas preelectorales se utiliza el muestreo probabilístico-aleatorio-estratificado¹⁷⁴, en el que

¹⁷⁰ En España, además del CIS (Centro de Investigaciones Sociológicas), algunas comunidades autónomas disponen de su propia agencia para la realización de barómetros de opinión temporales. También podemos encontrar universidades y equipos de investigación que las llevan a cabo sin la presión y subjetividad difusora posterior del “cliente” (ej. CEMOP Murcia).

¹⁷¹ “En Italia, la exactitud de las encuesta se ha visto menoscabada por la reticencia de quienes apoyan a los comunistas a revelar sus simpatías políticas a los entrevistadores.” (Kavanagh 1994:152)

¹⁷² El error que se comete debido al hecho de que se obtienen conclusiones del total de una población a partir del análisis de una parte de ella, se conoce como “error de muestreo”.

¹⁷³ Una muestra representativa de una población es el grupo, relativamente pequeño, de individuos de ese colectivo que presenta unas características semejantes a las de éste.

¹⁷⁴ El procedimiento de muestreo del CIS es polietápico, estratificado por conglomerados, con selección de las unidades priarias de muestreo (municipios) de la unidades secundarias (secciones) de forma aleatoria proporcional, y de las unidades últimas (individuos) por rutas aleatorias y cuotas de sexo y edad.

se divide a la población con capacidad de ejercer su derecho al voto en estratos y se escoge de forma aleatoria un número de individuos de cada uno de ellos proporcional¹⁷⁵ al número de componentes de cada estrato, y que presente unas características semejantes (franja de edad y sexo). El número de individuos que serán entrevistados finalmente se conoce como “tamaño de la muestra” y se obtiene mediante fórmulas de cálculo del tamaño muestral en las que intervienen una serie de parámetros particulares para cada estudio, como el tamaño de la población, la probabilidad de respuesta¹⁷⁶, el error de la estimación que se prevé cometer y el nivel de confianza. (AFHA, n.d., pp.1-3)

La representatividad de la muestra es, por tanto, y en relación al objeto de nuestra investigación, unos de los factores a tener en cuenta ya que lo que teóricamente hace que una muestra sea buena es la máxima aleatoriedad de la selección de los individuos que la componen (el hecho de que cada individuo tenga las misas posibilidades de formar parte de la muestra), en lo que supone que una de las razones por la que los estudios de predicción basados en el análisis de redes sociales podrían desestimarse o no tomarse como fiables, ha sido y es cuestionando si el conjunto de una población determinada (la población objeto de estudio: Toda la población electoral: españoles mayores de 18 años) se encuentra suficientemente representada en estas redes (aspecto mencionado en este capítulo en citas de Gayo-Avello, 2011; Willians y Gulati, 2009). Siendo así el estado de la cuestión, en el siguiente epígrafe, a través de varios informes sobre el uso de las Redes Sociales en España, vamos a tratar de extraer los datos necesarios para centrar la duda acerca de si es posible considerar que el conjunto de la población está representado en los usuarios de Redes, y/o qué estratos del colectivo podrían considerarse respaldados.

¹⁷⁵ El hecho de poder hacer conclusiones de una población a partir de una muestra se fundamenta en la inferencia estadística. Ésto se posibilita por la teoría de la probabilidad y en la ley de los grandes números.

¹⁷⁶ $P=Q=0,5$, para una situación de máxima incertidumbre en la respuesta.

3.5. Penetración de las Redes Sociales en la población española

Es sin duda estudiado y documentado el crecimiento generalizado en el acceso a Internet y el incremento del uso de la red, reduciéndose según avanza el tiempo, la llamada brecha digital en los países desarrollados¹⁷⁷. A nivel mundial y según datos de la Unión Internacional de Comunicaciones (UIT) en 2011 el número de usuarios de Internet en el mundo alcanzó los 2.000 millones¹⁷⁸.

En España, según los resultados de la “Encuesta sobre Equipamiento y Uso de Tecnologías de la Información y Comunicación en los hogares 2011” publicado por el INE (Instituto Nacional de Estadística), los datos más relevantes de cara a nuestra investigación, como son el acceso a Internet por franjas de edad y la participación en Redes Sociales, son los siguientes (INE, 2011):

- El 63,9% de los hogares españoles tiene acceso a la Red, frente al 59,1% del año anterior. En España existen 9,9 millones de viviendas familiares que tienen acceso a Internet, con un aumento de más de 800.000 hogares respecto a 2010.
- El 67,1% de la población de 16 a 74 años ha utilizado Internet en los tres últimos meses, lo que supone 23,2 millones de personas. El número de internautas ha crecido un 4,5% respecto al año 2010.

¹⁷⁷ Brecha digital hace referencia a la diferencia socioeconómica entre aquellas comunidades que tienen accesibilidad a Internet y aquellas que no, aunque tales desigualdades también se pueden referir a todas las nuevas tecnologías de la información y la comunicación (TIC), como el computador personal, la telefonía móvil, la banda ancha y otros dispositivos. Como tal, la brecha digital se basa en diferencias previas al acceso a las tecnologías.¹ Este término también hace referencia a las diferencias que hay entre grupos según su capacidad para utilizar las TIC de forma eficaz, debido a los distintos niveles de alfabetización y capacidad tecnológica. También se utiliza en ocasiones para señalar las diferencias entre aquellos grupos que tienen acceso a contenidos digitales de calidad y aquellos que no. El término opuesto que se emplea con más frecuencia es el de inclusión digital y el de inclusión digital genuina. (Maggio, 2007)

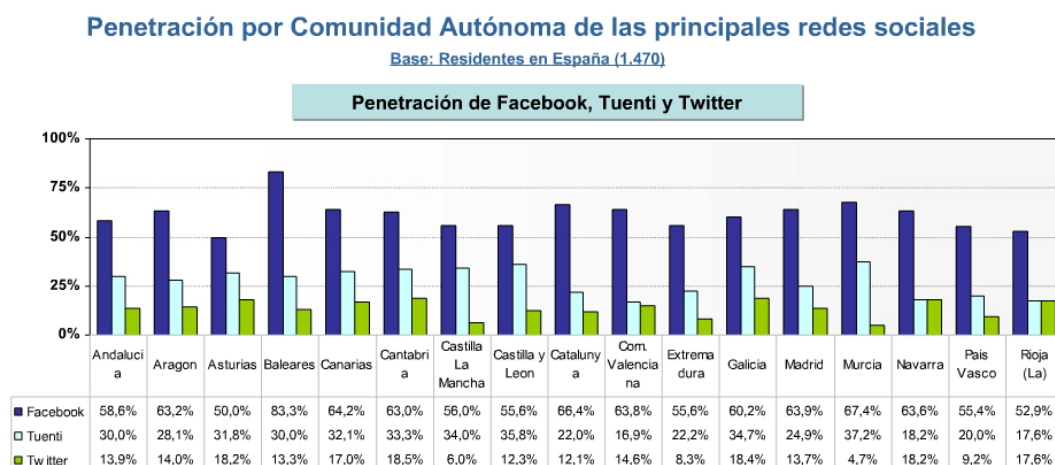
¹⁷⁸ "UIT: Número de usuarios de Internet en el mundo alcanza los 2.000 millones". *Noticias24*. Publicado el 26 de enero de 2011 <http://www.noticias24.com/tecnologia/noticia/4373/uit-numero-de-usuarios-de-internet-en-el-mundo-alcanza-los-2-000-millones/>. (Consultado el 14/04/2012)

- El 52,3% de los usuarios de Internet en los últimos tres meses participa en redes sociales de carácter general, como Facebook o Twitter, creando un perfil de usuario o enviando mensajes u otras contribuciones. Los más participativos son los estudiantes (90,2%) y los jóvenes de 16 a 24 años (88,5%). Por sexo, las mujeres participan más que los hombres (54,8% frente al 49,9%).
- Las comunidades autónomas que se encuentran por encima de la media española en los tres indicadores sobre el uso de ordenadores e Internet son Aragón, Illes Balears, Cantabria, Cataluña, Comunidad de Madrid, Comunidad Foral de Navarra y País Vasco.

Son datos destacables los referidos a la participación de la población en Redes Sociales como Facebook y Twitter ya que nos indican que el 52% de los 23 millones de personas entre 16 y 74 años (mayoritariamente población con derecho a voto) que utilizan Internet, participan en dichas redes. Además, se incluye la información acerca de las comunidades autónomas dado que dos de ellas, Comunidad de Madrid y Cataluña, han celebrado elecciones que se han monitorizado para su estudio.

Continuando con el grado de penetración de las Redes Sociales por Comunidad Autónoma, el “Estudio de Hábitos de Internet 2010” profundiza en conocer como se manifiesta esta penetración por comunidades. Estos datos que arrojan el porcentaje de la población de un área concreta que participa en Facebook y Twitter, podrán ser utilizados en nuestro análisis en aras de argumentar si dichos usuarios podrían o no considerarse como muestra más o menos representativa de la población en los procesos electorales monitorizados. No obstante, el mismo estudio advierte que su información debe tomarse con cautela y a modo de orientación puesto que en las comunidades más pequeñas la muestra es reducida, mientras que para Andalucía, Cataluña, Madrid y Valencia se cuenta con una muestra considerable. (Ocio Network, 2010)

Ilustración 24. Penetración de Redes Sociales por comunidad autónoma (España), realizado por Ocio Network (2009)



Un segundo estudio dedicado al desarrollo de la sociedad de la información en España, publicado por la Fundación Orange también manifiesta un marcado incremento en la pertenencia y participación a Redes Sociales como Facebook, Tuenti y Twitter: “*El Estudio de Hábitos de Internet 2010, elaborado por Ocio Networks, señala que el 83% de los internautas españoles utiliza algún tipo de red social. Facebook sigue siendo la más usada (tres de cada cuatro usuarios) y, además, creciendo a una tasa superior al 20%. Después de ella se posiciona Tuenti (uno de cada tres internautas), seguida de Twitter, la red que más crecimiento ha experimentado en el año 2010.*” (Fundación Orange, 2011, pp.146-147)

El informe, basándose en terceros estudios originados por otras consultoras, continua exponiendo: “*De acuerdo con el II Estudio sobre Redes Sociales en Internet, publicado por IAB Spain Research, estas redes se caracterizan por un lado, por la elevada frecuencia en que se visitan, pues en el caso concreto de Facebook y Tuenti, casi un 40% de usuarios de estas redes afirman visitar su red social varias veces al día... Cada vez son menos los internautas que no usan una red social, ya sea como medio de comunicación o por motivos profesionales, argumentando como principales razones para no usarlas la falta de interés, de tiempo, de privacidad o simplemente porque no son de su agrado. Pero lo que sí parece reflejarse es que el*

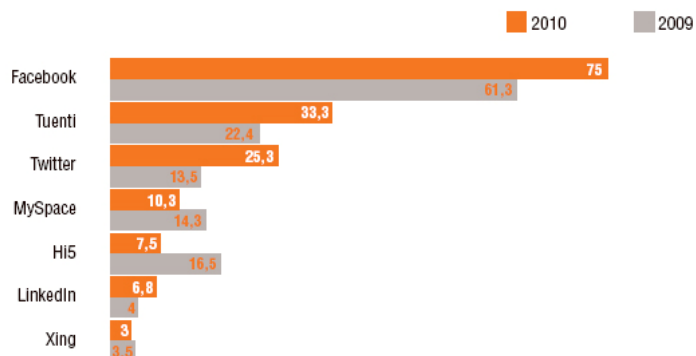
uso de estas herramientas es más popular entre la población más joven.” (Fundación Orange, 2011, pp.149-150)

Al igual que sucedía en la intención de conocer el porcentaje de usuarios de Redes Sociales por comunidad autónoma, el conocimiento del uso que se hace de éstas por franjas de edad es relevante para el objeto de nuestra investigación, pues nos permitirán con estos datos de nuevo, conocer si sus usuarios son representativos de la población de estudio en unas elecciones dadas. A partir de aquí, vamos a mostrar los datos revelados al respecto por varios estudios desarrollados en el territorio español.

Ilustración 25. Gráficos de uso de Redes Sociales en España realizado por Fundación Orange (2011)

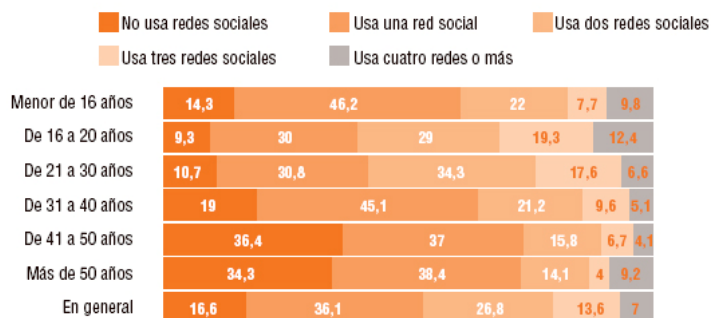
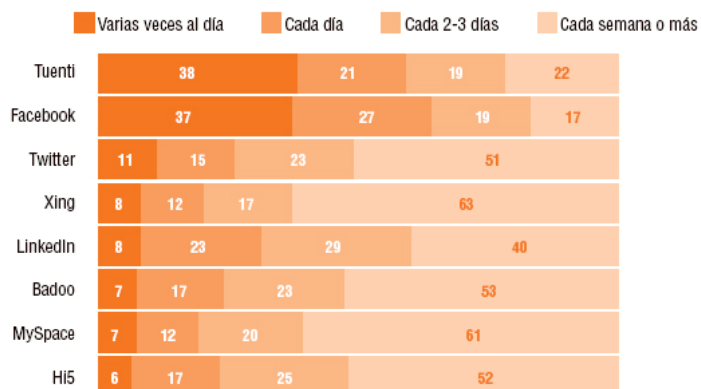
> Gráfico 6.27. Principales redes sociales utilizadas en España. 2009-2010, en % sobre el total de usuarios

Fuente: Ocio Networks (2011)



> Gráfico 6.28. Frecuencia de uso de las redes sociales en España. 2010, en % sobre el total de usuarios

Fuente: eEspaña 2011 a partir de IAB Spain Research (2010)



> Gráfico 6.31. Número de redes sociales usadas por los internautas. 2010, en % sobre el total de usuarios

Fuente: Ocio Networks (2011)

El 75% de usuarios conectados es usuario de Redes Sociales en el año 2011 según apunta el “III Estudio sobre Redes Sociales e Internet” realizado por Elogia e

IAB Spain Research¹⁷⁹. De ellos, el 61% son mujeres y el 39% hombres. Por franjas de edad: de 18 a 30 años y de 31 a 39 años, un 34%, y de 40 a 55 años, un 32%. Facebook, con un 95% de usuarios, es la red social más utilizada por los encuestados, y Twitter alcanza un 34%. En ambas, la edad es un elemento transversal, encontrándose entre un 92-98-96% en el caso de Facebook y entre un 37-31-33% en Twitter. En cuanto a intenciones de uso futuro, Twitter gana posiciones con respecto a Facebook, con un 37% frente a un 29% de entrevistados que considera que crecerá su uso en el próximo año. (Elogia-IAB, 2011)

Similares resultados se desprenden del trabajo publicado por “The Cocktail Analysis” (“4ª Oleada Observatorio de Redes Sociales Las marcas empiezan a encontrar límites”), según el cual Facebook¹⁸⁰ afianza su liderazgo con un 85% de usuarios y Twitter, que duplica su número respecto a 2010, alcanza un valor de 32%. Esta red, “aparece como la red de moda, gracias a la presencia de *celebrities* y la llegada masiva de jóvenes tras estos.” Por sexo, 53% hombres y 47% mujeres en Facebook, y 61% hombre y 39% mujeres en Twitter. Por franjas de edad, el 70% y el 56%, respectivamente, tiene más de 25 años¹⁸¹. Además, indican que ambas redes resultan atractivas para los más jóvenes por su presencia en los medios de comunicación. También se apunta que un 65% de los usuarios que mantienen un perfil en Facebook han mantenido alguna relación con una empresa o marca, y en Twitter la cifra alcanza el 32%. Si bien el estudio está centrado en marcas comerciales, esta última aportación es de interés de cara a nuestra investigación porque podríamos extrapolar que de igual modo que los usuarios de Redes Sociales emplean éstas para comunicarse con empresas, podrían hacerlo para mantener contacto con candidatos y partidos, si bien no disponemos a día de hoy de ese dato concreto y no sabemos qué porcentaje de usuarios de Redes Sociales mantienen una

¹⁷⁹ La muestra, de 966 casos, se refiere a los residentes en España de 18 a 55 años de edad, por lo que los mayores de 55 no estarían representados.

¹⁸⁰ 1.304 encuestas online entre las que seleccionaron 26 para una entrevista en profundidad

¹⁸¹ Internautas con edades comprendidas entre los 16 y los 45 años. Los mayores de 45 años no están representados en estos resultados.

relación con candidatos y/o partidos políticos. También es importante mencionar que Twitter ha ampliado su carácter de canal en el que mantenerse informado, llegando a sustituir en muchos casos el consumo de noticias en esta Red a la consulta de medios de comunicación tradicionales. (The Cocktail Analysis, 2012)

Refiriéndonos al uso concreto de Facebook en España, el número de usuarios de Facebook que asciende a 14.343.980, representa el 31% de la población total de España¹⁸². En cuanto al reparto por sexo entre sus usuarios, el 49 % son hombres y el 50% mujeres (el 1% no manifiesta su sexo). Por franjas de edad destaca la comprendida entre los 25 y 34 años y por ciudades, Madrid y Barcelona ocupan los primeros puestos de uso. (Cuentamelared, 2011)

Ilustración 26. Gráfico de usuarios de Facebook en España (2011), por edad y sexo, realizado por Cuentamelared (2011)



¹⁸² Los datos expuestos en este análisis de usuarios de Facebook son aproximados, habiendo sido tomados de la plataforma de ayuda para diseñar anuncios en facebook, <http://www.facebook.com/business>

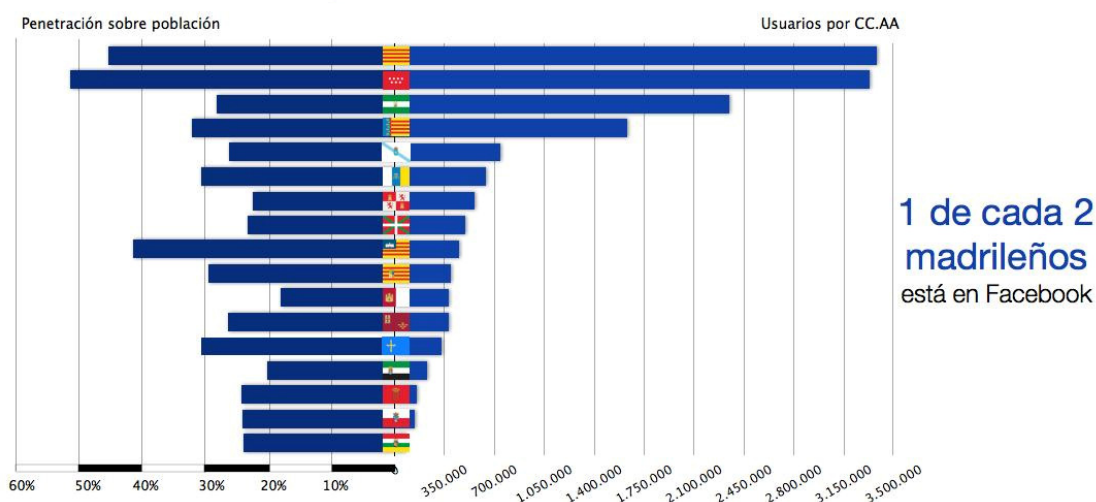
Más reciente, una infografía realizada por Zoom Labs en febrero de 2012 arrojaba la cantidad de 15,6 millones de usuarios Facebook en nuestro país¹⁸³. La mayor penetración por franjas de edad se da entre los usuarios con edades comprendidas entre los 15 y 29 años. La distribución por sexos se iguala al 50% de perfiles de hombres y mujeres. En cuanto a la penetración por Comunidades Autónomas, Cataluña y Madrid se sitúan en los primeros puestos con una penetración del 45% y más del 50% respectivamente. (Zoom Labs, 2012)

Ilustración 27. Gráficos de usuarios de Facebook en España, por edad y comunidad autónoma, realizado por Zoom Labs (2012)

Distribución por edad



Distribución por CC.AA



¹⁸³ El informe de la consultora SocialBakers arrojaba un dato similar: 16 millones de usuarios. <http://www.socialdente.com/2012/02/01/numeros-de-facebook-en-espana-enero-2012/>. (Consultado el 21/04/2012)

Según el estudio “Observatorio de las redes sociales, IV oleada”, el 85% de los internautas utilizan Facebook, siendo un 53% hombres y un 47% mujeres.

Para conocer las franjas de edad en 2012, nos acercamos a una infografía realizada por Michele Caivano (fuente de datos: Facebook AD) que recoge datos de los 15 países con mayor presencia en Facebook y entre los que se encuentra España con 15,7 millones de usuarios. El porcentaje de usuarios de género masculino es de 49,6% y de 50,4% para las mujeres.

- De 13-17 años: 8,4%
- De 18-24 años: 20,7%
- De 25-34 años: 31,7%
- De 35-44 años: 22,7%
- De 45-54 años: 10,8%
- De 55-64 años: 3,8%
- Más de 65 años: 2,1%

Ilustración 28. Gráfico de usuarios de Facebook en España, por sexo, realizado por Michele Caivano (2012)



Si nos fijamos en el uso de Twitter en España, a finales de 2010 el número de cuentas en esta red alcanzaba una cifra cercana a los 3 millones de usuarios¹⁸⁴. Por franjas de edad, según el informe “Observatorio de las Redes Sociales, 3ª oleada”:

¹⁸⁴ Semiocast ha analizado 383 millones de perfiles e Twitter creados antes del 1 de enero de 2012.

- De 16 a 18 años: 9%
- De 19 a 25 años: 30%
- De 26 a 35 años: 33%
- Más de 35 años: 29%

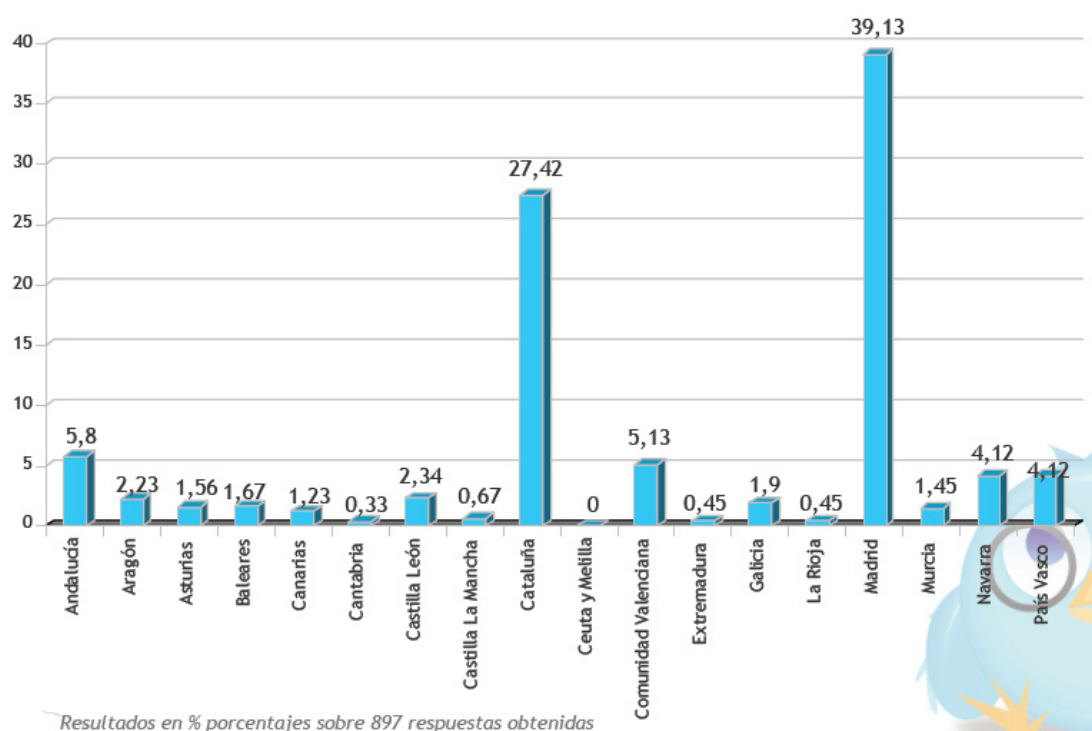
En el siguiente estudio del citado observatorio correspondiente a datos de 2011, en el que se afirma que el 32% de internautas hace uso de Twitter, los hombres están más presentes que las mujeres en esta red social: 61% a 39%. El 56% de sus usuarios tienen más de 25 años. (The Cocktail Analysis, 2012)

Según encuesta llevada a cabo por FECEMD, el uso de esta red por comunidades autónomas, destacan un 39,13% de usuarios en Madrid y un 27,42%, de Cataluña; a bastante distancia se encuentran Andalucía y Valencia con un 5,8% y 5,13%, respectivamente¹⁸⁵. Además, más de un 94% de los usuarios publica noticias, estudios y gráficos de interés profesional, así como páginas Web que considera interesantes. (Adigital, 2010)

La cifra de 3 millones de usuarios se ha más que duplicado en 2012, según un estudio a nivel mundial llevado a cabo por la agencia francesa Semiocast , logrando la cifra de 7 millones de cuentas en Twitter, siendo además el tercer país en cuanto a volumen de actividad se refiere. http://semiocast.com/publications/2012_01_31_Brazil_becomes_2nd_country_on_Twitter_superseds_Japan. (Consultado el 15/04/2012)

¹⁸⁵ 897 encuestas online contestadas satisfactoriamente en el mes de junio de 2010.

Ilustración 29. Gráfico de usuarios de Twitter en España, por comunidad autónoma, realizado por Adigital (2010)



Si bien es probado que el mayor uso de Twitter se da en los EEUU, España alcanza un loable noveno puesto en lo que a cuentas abiertas se refiere, y aún tanto o más importante, es el tercer país en volumen de actividad, lo que se traduce en que los españoles muestran una tendencia grande hacia la socialización en estas Redes y que son una población destacable en la participación en las mismas¹⁸⁶. Y participación supone difusión de mensajes, y esta difusión podría traducirse en amplificación del mensaje, de modo que unos pocos podrían influir en las convicciones de muchos al comentar los usuarios de esta red a otros individuos en espacios físicos lo que han leído en ella. Somos conscientes de que esta afirmación no es fácil de refutar con números, pero a la vez, estamos convencidos de que,

¹⁸⁶ De hecho, España es el tercer país en cuanto a volumen de actividad se refiere según estudio reciente de la agencia Semiocast.
http://semiocast.com/publications/2012_01_31_Brazil_becomes_2nd_country_on_Twitter_superseds_Japan. (Consultado el 15/04/2012)

siguiendo pautas demostradas de influencia, las personas influyentes en un círculo determinado, que contengan un alto nivel de educación y una alta participación en las Redes Sociales, podrían de esta forma influir en individuos alejados geográficamente o con los que no mantienen un contacto *face-to-face*¹⁸⁷.

Retomando la situación de Twitter en EEUU, país a la cabeza en el uso de esta red y que por tanto podría tomarse como país de referencia para tendencias de uso futuras en el resto de países que la manejan, el estudio “13% of online adults use Twitter” de *Pew Internet Research* evidencia que el uso de Twitter por usuarios comprendidos en la franja de edad entre los 30 y los 49 años se ha doblado entre 2010 y 2011. Este dato resulta de nuevo esperanzador de cara a la investigación que tratamos, y una vez más, aunque no se trate de datos cuantitativos de extrapolación directa al ámbito de estudio que se trata, sabida la realidad de que lo que acontece en los EEUU acostumbra a repetirse en el resto de países occidentalizados, nos anima a continuar en esta línea de investigación que aumenta significativamente la muestra de edad con capacidad representativa a la cercana edad de 50 años. (Smith, 2011)

Como puede deducirse siguiendo las cifras de penetración en las Redes Sociales, su uso es cada vez mayor entre la población española, alcanzando valores porcentuales muy altos entre aquellos que utilizan Internet. Esta consideración avala en cierto modo el objeto de nuestro estudio al poder argumentar que gran parte de la sociedad participa en ellas y, por tanto, de forma directa o indirecta, está expuesta a las opiniones ideológicas de sus “amigos” o “personas a las que siguen” en ellas. Sin duda un dato revelador lo encontramos en las diversas manifestaciones que “ignoran” a los usuarios con edades superiores a los 45-55 años (dependiendo de estudios)¹⁸⁸. Esta realidad supone que en nuestro estudio tengamos que considerar las alteraciones

¹⁸⁷ Dado que nuestro análisis es cuantitativo, estas consideraciones se incorporan exclusivamente como referencia y para consideraciones de futuras pautas en estudios venideros.

¹⁸⁸ Nos encontramos aquí con la “brecha digital” referida no a las diferencias entre países, sino como realidad consecuente con las fechas de adquisición del uso de las TIC. Las personas “mayores” demuestran dificultad –y quizás falta de interés por el esfuerzo que el uso de éstas les supone- para adaptar el uso de las nuevas tecnologías a su entorno vital.

que surjan de “desestimar este estrato de la población”, al menos, en las comparaciones que se establezcan con las encuestas preelectorales, en las que esta franja de edad encuentra su representación¹⁸⁹.

En el epígrafe dedicado a la exposición de la metodología se intentará establecer un análisis correlativo entre los usuarios de estas redes y la población con derecho a voto en nuestro país con el objeto de identificar qué estrato de la misma estaría representado en las Redes Sociales y cuál no, y así poder otorgar un rasgo de fiabilidad a los resultados obtenidos en el análisis de redes como herramienta de predicción electoral.

3.6. Conclusiones

Desde la campaña de Barack Obama en 2008, en la que su apuesta por Internet ha sido definida como el punto clave de su victoria electoral, derrotando primero a Hillary Clinton en las primarias y, posteriormente, a McCain en las generales, ha quedado más que patente la convicción de que cualquier candidato que quiera conseguir una mejoría en sus resultados o un éxito electoral, está de hecho obligado a desarrollar una cibercampaña.

Por otro lado, hemos visto como candidatos con mejores índices de popularidad en la Red no han visto reflejada tal ventaja en el resultado electoral obtenido, como es el caso de Antanas Mokcus, candidato a la presidencia de Colombia por el Partido Verde era el preferido en redes sociales como Facebook o Twitter y sin embargo fue derrotado por su oponente Juan Manuel Santos, consiguiendo, no obstante, ser la segunda fuerza política del país. Y también

¹⁸⁹ En menos de 30 años, dicha brecha digital habrá dejado de existir. Será entonces cuando el análisis de Redes Sociales y/o Internet se convertirá, casi con seguridad, en un canal de estudio de opinión extremadamente fiable. Véase como referencia a TAYLOR, Chris (2012). “Election 2040: The First President for Digital Natives”. *Mashable*. Internet. <http://mashable.com/2012/10/02/election-2040/> (Consultado el 05/10/2012)

podemos citar ejemplos en los que candidatos que no han desarrollado actividad web alguna, han resultado victoriosos en la contienda electoral.

Sin embargo, en los últimos años, los medios y redes sociales se han convertido en un espacio del mayor interés para la colaboración social y compartir información entre sus usuarios. La ingente cantidad de contenido que circula en ellas ha atraído la atención de agencias de marketing que esperan conocer el comportamiento de los usuarios para ajustar sus campañas online y de investigadores en el campo de las redes sociales que buscan en sus análisis revelar los modelos de difusión entre grupos y nodos en dichas redes, e incluso demostrar como ese contenido social puede ser utilizado para predecir resultados en el mundo real (Asur y Huberman, 2010). Recientemente este tipo de análisis ha saltado al campo de la política, y especialmente, al análisis de campañas electorales en la red.

Teniendo en cuenta los principales medios sociales en Red en los que hoy en día, al menos en España, se articulan las Ciber campañas Electorales, podemos contemplar varios espacios para el análisis del impacto de una campaña en red en Facebook, Twitter, Google y la página Web de partido y/o de candidato.

Además de Twitter, Red Social destacada por la facilidad de acceso a la información y las herramientas de análisis desarrolladas y al alcance, también podemos encontrar estudios que relacionan el número de seguidores o fans de una página Facebook de un candidato con su resultado electoral, y otros que han utilizado el volumen de búsquedas realizadas en Google sobre una opción política determinada.

Debido a los positivos resultados obtenidos en estos primeros estudios se ha apelado a la idea de que lo que la gente está publicando en su blog, o buscando en Google o difundiendo en Facebook o Twitter, proporciona un reflejo del sentimiento colectivo, lo cual puede traducirse en tendencia hacia un punto determinado. En este sentido, recientes estudios han anunciado que es posible predecir un resultado electoral mediante el análisis de este contenido social.

Concretamente, Tumasjan et al., (2010), sobre las elecciones de 2009 en Alemania, concluyeron que Twitter puede ser utilizado como predictor del resultado electoral dado que el mero número de *tweets* que mencionan a un partido político puede considerarse un reflejo posible de la cuota de voto y por tanto, su poder de predicción se acerca a las encuestas electorales tradicionales.

O'Connor et al. (2010) relacionaron la opinión pública en EEUU en el periodo 2008-2009 medida en las encuestas tradicionales con el sentimiento reflejado en los tweets (*sentiment analysis*), obteniendo datos dispares. Sus resultados eran muy coincidentes en algunos casos (con una correlación del 80% en la mayoría) y menos en otros: para *presidential job approval* en 2009 los datos de Twitter replican los obtenidos en las encuestas y para *pesidential elections polls* in 2008 la correlación no fue tan significativa. Los autores sugieren que es necesario mejorar las herramientas de NLP (programación neurolingüística) para mejorar dichas estimaciones.

Tal y como hemos comentado antes, las características de privacidad de la red social Facebook dificultan la labor de análisis de interacción en la red. Si bien en el análisis de datos en Twitter hay que tener en cuenta o habría que tener en cuenta variables externas a la propia red, en Facebook, esta realidad es una obligación de entrada. Los análisis publicados hasta la fecha, siempre incorporan una serie bastante numerosa de variables a cada candidato, dado que en Facebook, salvo los perfiles de los propios candidatos y partidos (con datos como número de “like” o “supporter” o “seguidores”, comentarios y publicaciones) y su evolución en pre y campaña, es prácticamente el único dato que puede obtenerse de manera automática por motivos derivados de las condiciones de privacidad de los usuarios de esta Red.

En varios trabajos de Willians y Gulati (2008, 2009), centrados en el análisis de Facebook en campaña electoral en EEUU, se concluye que el número de Fans constituye un indicador de la viabilidad del candidato y, por tanto, del éxito electoral. En su análisis incorporan una serie de variables para cada candidato, como su posición de peso en el partido, el hecho de que acudan a su reelección, que se

encuentren en una campaña competitiva y las características de su electorado (alto nivel de educación, blancos, residentes en núcleos urbanos y jóvenes, son más propensos a participar en Facebook).

Lui, Metaxas y Mustafaraj (2011) en su análisis en Google Trends en las elecciones de 2008 y 2010 en EEUU, afirman que Google Trends no es un buen predictor en ninguno de los dos procesos electorales analizados. Los autores argumentan que las causas de estas desviaciones pueden deberse por un lado al hecho de cuando un usuario busca por un candidato y no por su oponente, el resultado es que este último no alcanza los niveles mínimos de curiosidad y/visibilidad (permaneciendo en el vacío) lo que repercute más positivamente en el candidato buscado; por otro, que los electores pueden buscar por un candidato pero con el objetivo de encontrar información negativa lo que provoca ese margen de error con los resultados electorales; y, finalmente, que aquellos candidatos que ya ocupan un cargo político sobresaliente pueden ser buscados a causa de esta actividad y no tratando de encontrar información sobre su campaña.

En los estudios realizados sobre el impacto de las páginas Web's de partidos o candidatos, al igual que ocurría con el análisis de Facebook, se incorporan una serie de variables relacionadas directamente con el “valor” o “posición” del candidato en la campaña.

D'Alessio¹⁹⁰ en su análisis de las elecciones al Congreso de EEUU en 1996 ya concluía que contar con una cibercampaña proveía al candidato con 9,3 más votos de los previstos inicialmente, valores únicamente superados por la variable de la incumbencia, esto es, presentarse a la reelección, y por la posición más o menos relevante que ocupa dentro de su partido. Más recientemente, Gibson y McAllister (2005) estudiaron el efecto de las cibercampañas en el resultado electoral en las elecciones australianas de 2004, encontrando similares resultados: una

¹⁹⁰D'ALESSIO, David W. (1997). “Use of the Web in the 1996 US Election”. *Electoral Studies* 16(4): 489-501. Mencionado por Gibson y McAllister (2005).

cibercamapaña es un elemento primordial a la hora de asegurar la victoria electoral. El uso de una página Web proporciona un 4% adicional a la preferencia de voto inicial, lo cual supone más que todos los métodos tradicionales de campaña, excepto la incumbencia.

Se pudo observar por lo concluido hasta ahora que el análisis de redes sociales como método de predicción electoral ha obtenido resultados bastante dispares. Así, mientras que siguiendo los métodos tradicionales de encuestas, el margen de error no supera o debe superar el 2 o 3% (es preciso tener aquí en cuenta la notable desviación mencionada en este capítulo, con referencia al caso de las encuestas de las Elecciones Autonómicas Andaluzas 2012, en las que este margen se superó), en este tipo de análisis el margen de error puede llegar a alcanzar, e incluso superar en limitados casos, el 10%, si bien es cierto que en algunas ocasiones el margen de error conseguido es incluso menor que el publicado por las encuestas.

Daniel Gayo (2010 y 2011), advierte en su análisis electoral en Twitter, que las Redes Sociales no pueden ser todas como herramientas fiables de predicción y que aquel que las utilice como tales, cometería un grave error. En sus estudios, no obtuvo coincidencias suficientes para afirmar que Twitter pueda ser empleado como herramienta de predicción electoral.

La predicción de mercados (basados en apuestas) han resultado ser una nueva fuente de pronóstico electoral. Varios estudios académicos han confirmado el valor de los mercados de predicción en el pronóstico de próximos eventos, para los resultados electorales, demostrando su eficacia desde las elecciones de finales del XIX y las primeras elecciones del siglo XX. David Rothschild exploró la precisión de las previsiones derivadas de dos tipos diferentes de datos: las encuestas y los mercados de predicción. Las conclusiones de su estudio confirman los resultados logrados por estudios similares, como el de Rhode y Strumpf (2004) en el que afirman que los mercados de predicción pueden conseguir una predicción más exacta que la elaborada por las encuestas de opinión tradicionales.

Con similar metodología, Nate Silver un estadístico del béisbol que se inició el análisis de las encuestas políticas, presentó en marzo de 2008 su sitio *FiveThirtyEight.com*, en el que utilizaba su propia fórmula para predecir los resultados federales y estatales de las elecciones presidenciales de 2008 en EEUU, consolidando la victoria de Barack Obama.

De hecho nuestra conclusión es que no siempre es posible predecir elecciones a partir de los datos extraídos de la participación en Redes Sociales. Por una parte, los resultados obtenidos en un proceso, no son directamente extrapolables a otro, tanto porque los sistemas electorales son diferentes como porque los espacios en los que se producen presentan características demográficas bien diferentes. Por otra, los métodos de correlación que se emplean son también diferentes, siendo de unos casos los datos comparados con resultados electorales y en otros con encuestas preelectorales. Y por último, las herramientas y métodos de análisis, novedosos hasta el momento, deben ser testados, y continuamente contrastados y mejorados para asegurar el veraz resultado de los datos obtenidos. Además, hemos de tener en cuenta que no todas las franjas de edad son usuarios activos en estas Redes Sociales, siendo los mayores de 50 años los menos propensos a participar y compartir así sus opiniones y tendencias de voto, dada su bajo nivel de presencia y actividad en estos canales.

No obstante, se abre un campo de especial interés y relevancia de cara al futuro, en el que tal vez llegue a ser posible sustituir a las encuestas, que hasta hoy han sido una fuente fiable de la tendencia electoral.

Capítulo 3

Metodología

4. Capítulo III: Metodología

El análisis que se ha realizado sobre cuatro campañas electorales en España, acontecidas en los años 2010 y 2011, se refiere al análisis de los datos capturados en la Red Social Twitter durante el periodo de campaña electoral (15 jornadas previas a la jornada electoral).

Genéricamente, se ha dividido en dos apartados principales, que explican el método empleado para la fijación de la metodología y la rutina utilizada en la captura de datos en esta Red.

4.1. Introducción

El objetivo de este trabajo es conseguir, a través de la recopilación de los datos extraídos de Twitter durante el periodo de campaña, una conclusión en base a análisis cuantitativos, que pueda afirmarnos o desdecirnos acerca de la hipótesis de si es posible utilizar el análisis de Redes Sociales para predecir resultados electorales del mismo modo en que se viene haciendo tradicionalmente por las encuestas de opinión.

Segundas hipótesis podrían ser desveladas, como la que hace referencia a que las campañas electorales afectan a los individuos en su decisión de voto (Gelman y King 1994¹⁹¹, Crespo et al. (2004), y por tanto afectarían directamente al resultado electoral, y si los factores “tamaño de la muestra” y “fecha de realización de la encuesta” influyen directamente en los márgenes de error de las encuestas (Lau, 1994; DeSart y Holbrook, 2003).

En el campo de la ciencia política destaca el análisis de datos en Twitter tanto por ser una red extendida y reconocida como medio de expresión y difusión de

¹⁹¹ GELMAN, Andrew y KING, Gary (1994). "Party Competition and Media Messages in U.S. Presidential Elections." En *The Parties Respond* 2ª edición, editorial L. Sandy Maisel.

opiniones, como por la facilidad que ofrece para capturar dichos datos gracias a su API de libre disposición y explotación que permite recolectar lo que se conoce como *public time-line*, esto es todos los *tweets* de sus usuarios salvo los de aquellos que mantienen un perfil privado que protege sus publicaciones del resto. Las características de privacidad de la red social Facebook dificultan la labor de análisis de interacción en la red. Si bien en el análisis de datos en Twitter hay o habría que tener en cuenta variables externas a la propia red, en Facebook, esta realidad es una obligación de entrada. Los análisis publicados hasta la fecha, siempre incorporan una serie bastante numerosa de variables de control para cada candidato, dado que en Facebook, salvo los perfiles de los propios candidatos y partidos (con datos como número de “like” o “supporter” o seguidores, y comentarios y publicaciones) y su evolución en pre y campaña, es prácticamente el único dato que puede obtenerse de manera automática por motivos derivados de las condiciones de privacidad de los usuarios de esta Red. Es decir, en Facebook es posible obtener los datos de la actividad existente en la página de perfil o página fan de una opción política, estando en los límites de ella misma los límites de la obtención de datos. Por tanto, en esta red social, la actividad de los usuarios fuera de dichas páginas, no puede ser fácilmente monitorizada.

Es por esto y por el hecho de que, dadas las propiedades de Twitter, existen antecedentes investigadores en el terreno de procesos electorales y mercados de opinión, que para la investigación se centra en el análisis de los datos recogidos en Twitter durante cuatro procesos electorales acontecidos en España durante los años 2010 y 2011:

- Elecciones al Parlamento de Cataluña 2010
- Elecciones Autonómicas de la Comunidad de Madrid 2011
- Elecciones Autonómicas de la Región de Murcia 2011
- Elecciones Generales 2011

La selección de estos cuatro procesos electorales, tal y como se describió en la introducción de este trabajo, viene motivada por las siguientes razones:

- La novedad de la propia Red Social (Twitter), ya que antes del año 2010 su uso en España era bastante reducido en lo que se refiere a número de participantes en la misma, lo que supondría que la sociedad española apenas se encontraría representada.
- Las primeras investigaciones basadas en el análisis de datos en Twitter datan del año 2010, por lo que es a partir del conocimiento de su posibilidad cuando iniciamos la captura de datos del primer proceso electoral referenciado.
- Se monitorizan las elecciones que tuvieron lugar en el transcurso de un año. Ampliar dicho periodo dificultaría el establecimiento de elementos comunes de comparación entre los objetos observados, dado que entendemos que la población de estudio soportaría cambios estructurales relevantes como un incremento significativo del número de usuarios Twitter.
- Al tratar de procesos electorales que suceden en distintos ámbitos geográficos, siendo uno de ellos de carácter nacional y el resto de ámbito autonómico o regional, los resultados de sus análisis podrán arrojar conclusiones que indiquen las posibles diferencias entre el comportamiento Red en una elección atendiendo a las particularidades del entorno en que ésta se produce.

Fijándonos en métodos utilizados en varios de los trabajos mencionados a lo largo de esta obra, se ha decidido seguir como base inicial la técnica empleada en el publicado por Tumasjan et al., (2010) en el que sus autores concluyen que el número de *tweets* en los que se menciona (conocido como “menciones”) a uno u otro opción política puede considerarse como un reflejo válido y previsible de la intención de voto, y por ende, del resultado electoral, acercándose a los métodos predictivos obtenidos por las encuestas tradicionales.

Se ha desestimado el “análisis sentimental” o *sentiment analysis*, y el “análisis de influencia”, dado que su estudio no es objeto de esta investigación. La identificación de usuarios influyentes no afectaría a nuestros resultados finales, los cuáles no toman en cuenta difusiones particulares y el modo en que éstas se diseminan, sino el volumen de participación en su totalidad. Tampoco perseguimos determinar el carácter cualitativo positivo, negativo o neutro, de dichos mensajes en cada caso monitorizado. Además, existen segundas razones que se expresan a continuación.

En el caso de los análisis de influencia, es preciso mencionar que aquellos de carácter investigador y/o académico (Asur et al., 2011; Zhu, Huberman y Luon, 2012), más allá de aplicaciones comerciales como *Twentyfeet* -que monitorea de forma aislada y no en el entorno de una red dada, el número de seguidores, número de menciones y listas a las que pertenece un usuario determinado para calcular su grado de influencia-, son escasos y emplean fórmulas de cálculo complejas que no refieren detalladamente en sus publicaciones, dificultando e incluso imposibilitando, su replicación exacta.

En el caso del *sentiment analysis*, es la constatación por Asur y Huberman (2010) de que este tipo de análisis no incorpora mejoras en comparación con el análisis único de los tweets: “*We find that the sentiments do provide improvements, although they are not as important as the rate of tweets themselves.*” (Asur y Huberman, 2010, p.7). Por otro lado y contrariamente a lo que sucede en otros ámbitos, mientras que se venía considerando que los adjetivos afectivos cumplían un papel definitivo en la clasificación de los sentimientos, en el dominio de la política, estas palabras no son los mayores indicadores del tipo de la información que se vierte, siendo los nombres (de partidos y candidatos) modos de expresión política válidos en sí mismos: “*Furthermore, affective adjectives are not the most informative political opinion indicators. Instead, the choice of topics, as reflected in neutral nouns, is an important mode of political opinion expression by itself. Our*

manual annotation results demonstrate that a significant number of political opinions are expressed in neutral tones.” (Yu et al., 2008, p.9).

4.2. Metodología para el análisis

4.2.1. Primera línea de análisis: toma de datos en la campaña electoral

Es necesario anotar que respecto al seguimiento fiel del método empleado por Tumasjan et al., (2010) en las elecciones alemanas, en las que se toman como referencia únicamente las menciones a los partidos, para la investigación que se expone se han tomado en cuenta la presencia y difusión tanto de partidos como de candidatos. Además, dichas menciones se contabilizarán de forma única y en análisis paralelos, junto con el resto de actividad Twitter, como Followers, RT's, tweets publicados y Reply's. Cada una de ellas se ha definido como variable independiente.

La captura de los *tweets* se ha realizado empleando el API de Twitter. Se han descargado con total continuidad, disponiendo de servidores y otros medios técnicos destinados a evitar rupturas o paradas del servicio que hubieran provocado, en caso de producirse, pérdidas de información.

La *query*¹⁹², o consulta, lanzada para la búsqueda contiene los nombres de los perfiles de candidatos y partidos (en sus tres posibles variantes: texto, @ y #), y los *hashtags* de la campaña. Con estos datos de consulta no aseguramos de capturar todos los tweets que hablan sobre dicha campaña. Con el fin de minimizar la duplicación de perfiles o el seguimiento de perfiles falsos, los datos se han obtenido de las páginas Web's oficiales de cada opción política (Zarrella, 2010). En el mismo sentido, Lui, Metaxas y Mustafaraj (2011) en su estudio de predicción electoral sobre el volumen de búsquedas reflejadas en Google Trends, determinaron la “oficialidad”

¹⁹² Las consultas son el mecanismo principal para recuperar información de una base de datos y se componen de preguntas presentadas a la base de datos en un formato predefinido.

de los nombres de los candidatos tal y como debieran buscarse siguiendo la lista que presentaba el diario The New York Times: “*We also downloaded the names of the candidates and the results of elections using the New York Times (<http://www.nytimes.com/politics/>), one of the more authoritative newspapers in the US, which maintains detailed data of prior and current elections, including information of incumbent success.*” (Lui et al., 2011, p.3)

No obstante, en algunos casos se han incorporado *hashtags* creados por los propios usuarios y que se han considerado por su aceptación en la red y, por tanto, por su capacidad de difusión.

El periodo de captura para todos los casos analizados se ha establecido en los quince días de campaña previos al día de celebración de cada una de los procesos electorales monitorizados. Además, todas las medidas se han realizado por días para que puedan ser agrupadas de una forma más flexible y en caso de necesidad, analizadas individualmente.

Para la determinación del mismo se han seguido los patrones empleados por Tumasjan et al., (2010) y Gayo-Avello (2001), ambos considerados como investigaciones de referencia e inspiración de nuestro trabajo. Los periodos de captura próximos a un determinado evento fueron también seguidos fuera del análisis electoral. Asur, y Huberman (2010) definieron el *critical period* con una duración de tres semanas (la previa al lanzamiento y dos posteriores): “*For each movie, we define the critical period as the time from the week before it is released, when the promotional campaigns are in full swing, to two weeks after release, when its initial popularity fades and opinions from people have been disseminated.*” (Asur y Huberman, 2010, p.2)

Según la metodología de la ciencia política, las hipótesis explicativas incluyen, al menos, una variable dependiente y una independiente. En el caso que ocupa, la variable dependiente es el “Resultado Electoral” y como variables

independientes¹⁹³ se han definido por separado cada uno de los tipos de datos obtenidos (que se corresponden con un actividad concreta en Twitter, relacionada con la actividad y seguimiento de líderes y partidos políticos)¹⁹⁴.

Si bien “... *se trata de intentar explicar lo más posible con el menor número de variables* (King, Keoane y Verba, 194, p.29)” (Anduiza, Crespo y Méndez, 2009) el investigador es libre de incluir cuantas variables considere para mejorar su método. Nótese que si bien se enumeran 11 variables independientes, éstas aparecen “duplicadas”, para partido y candidato. Es como si se tratara de 5 (más una sexta dedicada a los hashtag de campaña) que permiten su cálculo unido (candidato más partido) y separarse en sendos grupos (candidato y partido) para ampliar conclusiones.

Los partidos políticos (y sus respectivos candidatos) que se han tenido en cuenta en el estudio son, en todos los casos, aquellos de carácter nacional que han obtenido mayor tasa de representatividad, desestimándose partidos minoritarios y aquellos de carácter regional.

El total de las variables independientes definidas y analizadas en todos los casos monitorizados, son:

Tabla 12. Relación de Variables Independientes

Nomenclatura	Tipo de variables independientes
V1	Followers candidato
V2	Followers partido
V3	Tweets publicados por candidato

¹⁹³ Las categorías “candidato” y “partido” se han mantenido separadas en cada caso, para maximizar y ampliar posibles conclusiones.

¹⁹⁴ No se ha incluido finalmente la variable “Global Clicks”, referida al número de veces en que la página URL de un partido o candidato ha sido citada en Twitter porque sólo pudo obtenerse en dos de los procesos analizados.

V4	Tweets publicados por partido
V5	Menciones a candidato (en @, en # y en texto)
V6	Menciones a partido (en @, en # y texto)
V7	Menciones a hashtags de campaña
V8	RT a candidatos
V9	RT a partidos
V10	Reply a candidatos
V11	Reply a partidos

Fuente: elaboración propia

Si bien es cierto que para un mayor control y veracidad del resultado del análisis debieran haberse tenido en cuenta en cuenta variables relacionadas directamente con las características de la campaña y de los candidatos, como son la posición de peso de estos en el partido, el hecho de que acudan a su reelección, las particularidades de su electorado, o que se encuentren en una campaña competitiva. Aún conscientes de su relevancia y debido a la carencia de dicha información, se ha prescindido de ellas en el presente trabajo.

Se ha realizado una distribución proporcional al porcentaje de votos, que aunque limita metodológicamente el análisis, permite equiparar ambos datos (resultados electorales-actividad Twitter). Las correlaciones entre la variable dependiente (resultado electoral) y las variables independientes (datos de actividad en Twitter clasificados por candidatos y partidos), se establecen totalizando¹⁹⁵ la suma de los valores de las variables independientes desde el día del inicio de campaña hasta el día previo a la propia jornada electoral, y por la parte de la variable dependiente, los porcentajes de los resultados electorales definitivos publicados en las páginas oficiales de las instituciones competentes en cada proceso electoral

¹⁹⁵ En el caso de *followers* y *tweets*, se ha restado al valor de la fecha final, el valor de la fecha inicial.

vigilado. Los valores obtenidos de las sumas de los valores diarios de todas las variables independientes se han “totalizado” de forma que su 100% se ha correspondido con el porcentaje final resultado de la suma de los porcentajes parciales de los resultados electorales de las opciones políticas incluidas. De esta forma, se ha tenido en cuenta en la investigación el hecho de que ha existido una actividad Twitter sobre los partidos y candidatos no incluidos en la muestra, no habiéndose capturado el 100% de la campaña en esta Red, sino el porcentaje que corresponde al porcentaje de votos obtenido por los partidos y candidatos monitorizados.

Direcciones Web utilizadas para obtención de resultados electorales definitivos	
Elecciones al Parlamento de Cataluña 2010	http://www.gencat.cat/governacio/eleccions/eleccions2010/resultats2010/09AU/DAU09999CM_L2.htm
Elecciones Autonómicas a la Comunidad de Madrid 2011	http://resultados2011.madrid.org/12AU/DAU12999CM_L1.htm
Elecciones Autonómicas a la Región de Murcia 2011	http://resultados elecciones2011.carm.es/15AU/DAU15999CM_L1.htm
Elecciones Generales 2011	http://elecciones.mir.es/resultadosgenerales2011/99CG/DCG99999TO_L1.htm

4.2.2. Segunda línea de análisis: estimación de la ponderación de las variables

Se ha diseñado un método simultáneo comparativo cuyo objetivo ha sido tratar de conseguir un peso específico para cada una de las variables independientes

en el conjunto de la estrategia de campaña articulada en Twitter, determinando así el orden de relevancia que supuestamente éstas han ocupado en la misma. Para ello, se ha asignado a cada una de las variables un valor de ponderación, en una escala de 1 a 10.

Tabla 13. Valor de Ponderación asignada a cada variable

Ponderación	Variable
1	Followers candidato
1	Followers partido
1	Tweets publicados por candidato
1	Tweets publicado por partido
3	Menciones a candidato (en @, en # y en texto)
5	Menciones a partido (en @, en # y en texto)
9	Menciones a hashtags de campaña
4	RT a candidatos
6	RT a partidos
2	Reply a candidatos
4	Reply a partidos

Fuente: elaboración propia

Si bien podría argumentarse que los valores de ponderación incluyen un grado de subjetividad de aquel que los asigna, estos se han establecido teniendo en cuenta conclusiones de estudios publicados por diferentes autores referenciados en el marco teórico de esta investigación, y que a continuación se explican.

El bajo valor asignado a “*Followers*”, tanto para candidato como para partido se basa en la idea, atestiguada por diversos trabajos anteriores, según la cual este indicador supondría únicamente una muestra de “popularidad” y no de influencia, y además, las “celebrities” son objeto de seguimiento *per se*:

- Cha et al. (2010) afirman que el número de seguidores es un signo que revela la popularidad de un usuario (fuentes de noticias y figuras públicas) pero que no está relacionado con su capacidad de influencia entre la audiencia del mismo modo en que sí lo están menciones y retweets: “users who have high indegree do not necessarily spawn many retweets or mentions. This finding suggests that indegree alone reveals very little about the influence of a user.” (Cha et al., 2010, p. 8).
- Asur et al. (2011, p.7) corroboran tal afirmación: “The absence of correlation indicates that the number of followers is not an indication of influence, similar to observations in earlier work [9].”
- Y en nuestro ámbito geográfico, la pasión y por tanto sobrevaloración de las *celebrities* (entre las que podríamos incluir a los candidatos) se reconocen en la encuesta de The Cocktail Analysis (2012), según la cual Twitter aparece como la red de moda, gracias a la presencia de celebrities y la llegada masiva de jóvenes tras estos.

Es debido a esta característica, que tras realizar diversas simulaciones con distintos valores en la ponderación¹⁹⁶, se ha llegado a la conclusión de que el seguimiento, comentarios y aceptación de los candidatos en Twitter presenta una sobrevaloración en la aceptación popular en relación a la presencia de los partidos. Dicha vía de análisis encontró su fuente de inspiración y veracidad, además de en las anteriores consideraciones acerca de la tendencia convertirse en *Follower*, en el trabajo de Kavanagh (1994) en el que se observa una disparidad entre la popularidad del líder y la del partido a la que representa: “*En sistemas presidenciales, las preguntas claves habrán de centrarse en las personalidades en pugna y en la fuerza y debilidad de los candidatos. En contraposición, las elecciones parlamentarias en Gran Bretaña y Alemania Occidental han demostrado que la correlación entre la popularidad de un líder y el apoyo a su partido puede ser relativamente baja.*” (Kavanagh, 1994, p.6)

¹⁹⁶ Dichas conclusiones se demostrarán en el capítulo dedicado a los resultados de la investigación.

Respecto a “*tweets* publicados por partidos o candidatos”, entendemos que queda claro que son publicaciones efectuadas por los mismos interesados, y que si bien es necesario tenerlos en cuenta, únicamente podrían demostrar la actividad difusora y con un carácter publicitario de los propios sujetos, y en ningún caso, indicarían la participación o preferencia de los usuarios (posibles votantes) hacia una u otra opción política.

La siguiente variable, “Menciones a candidato/partido (en @, en # y en texto)” hace referencia a las menciones que se publican en los tweets acerca de partidos y candidatos, bien en forma de nombre de perfil (@), hashtag (#) o mención en texto (caracteres). Dicha variable por sí sola es el objeto de la investigación llevada a cabo por Tumasjan et al. (2010) acerca de las elecciones alemanas celebradas en el mes de septiembre de 2009. Si bien se tiene en cuenta los positivos resultados obtenidos en dicha investigación, se les ha otorgado un valor medio, dado que se ha considerado que *Hashtags* y *RTs* merecen una atención superior debida a lo que suponen sus características en la actividad Twitter, como podrá comprobarse a continuación. El valor del hashtag está también sedimentado en que los propios desarrolladores de Twitter basan en él su relación diaria de Trending Topics o TT, esto es, ¿de qué se habla ahora en el mundo?

En los trabajos ya mencionados de Cha et al. (2010) y de Asur et al. (2011), se manifestaba el poderoso valor de la siguiente variable, el “RT” o *Retweet*, considerándose esta variable por ambas investigaciones como uno de los mayores valores a tener en cuenta en cualquier proceso de comunicación en esta red. El RT, de hecho, es tomado en cuenta por muchas de las herramientas gratuitas disponibles en Internet para valorar la influencia o el peso de un individuo en Twitter. El RT supone que aquellos que nos siguen consideran que la información que publicamos merece tanto la pena como para difundirla al resto de nuestra red. Es por tanto que se hace RT de las publicaciones o *tweets* que se consideran merecen ser conocidas por nuestra red. La información que contienen es relevante y cuantos más RTs consigue un tweet determinado más valor adquiere la información que contiene. Por tanto,

cualquier información acerca de la campaña (eventos del programa o ideas de la opción política) que sea susceptible de retransmitirse a otros, adquiere un gran valor. No se acostumbra a hacer RT de contenidos superficiales (hacerlo supone que los que nos siguen podrían pensar que carecemos de profundidad de pensamiento y a lo largo, dejarían de seguirnos, por nuestro bajo perfil de interés o afinidad política). El ReTweet ha sido concebido desde el inicio como una forma de contagiar, difundiendo, nuevas ideas: “ReTweets may seem like a small idea, and they are. But that small idea is the first real window into how ideas spread from person to person.” (Zarrella 2009)

La mayor o menor primacía asignada a estas variables encuentra refutación en el estudio de la polarización en Twitter desempeñado por Conover et al. (2011): *“Using a combination of network clustering algorithms and manually-annotated data we demonstrate that the network of political retweets exhibits a highly segregated partisan structure, with extremely limited connectivity between left- and right-leaning users. Surprisingly this is not the case for the user-to-user mention network, which is dominated by a single politically heterogeneous cluster of users in which ideologically-opposed individuals interact at a much higher rate compared to the network of retweets.”* Sus palabras son aquí interpretadas, con independencia de la polarización de la red que no es objeto de este estudio, como que el RT mostraría un voto muy decidido mientras que las menciones podrían interpretarse como conversaciones entre votantes indecisos o discusiones entre votantes ya decididos. Se tiene en cuenta aquí, según conclusiones del estudio de Desart y Holbrook (2003), que los votantes indecisos no son determinantes en el nivel de acierto de una encuesta electoral.

Por último, la variable Reply. El Reply supone conversación con la opción política, en sentido positivo o negativo. Si bien a priori podría concedérsele un valor superior entendiendo que aquellas actividades Twitter destinadas a conversar con los políticos podrían ser tomadas de forma positiva e incrementadora, también es cierto que muchos de los Replys se deben a sensaciones contrarias, esto es, a quejas o

malas formas empleadas por los contrarios a una de las opciones políticas en contienda¹⁹⁷, por lo que este tipo de actividad no supone ser un indicador de tendencia de voto en todos los casos.

4.2.3. Propuesta de tercera línea de análisis: diferentes periodos en la toma de datos

Siguiendo resultados de las investigaciones de Desart y Holbrook (2003) y Lau (1994) dedicadas a la definición de las fuentes de error en la precisión de las encuestas preelectorales, se advierte que los autores concluyeron que la cercanía o lejanía de la fecha en que se realizaban las encuestas con la fecha de la celebración electoral era un factor influyente en el margen de error de éstas.

En teoría, los resultados de la encuesta ofrecen una estimación del parámetro de la población en el momento de la encuesta. Si el parámetro de la población cambia con el tiempo, entonces se debe esperar que las encuestas tomadas anteriormente no conseguirán una estimación tan precisa de los parámetros, como las tomadas más tarde. Asumiendo que una de las consecuencias de las campañas electorales es la capacidad de modificar la opinión (Holbrook, 1996), y que los votantes tomen conciencia de sus preferencias según progresa la campaña (Gelman y King, 1994), se espera que las encuestas tomadas más cerca al día de las elecciones proporcionarán mejores estimaciones que los resultados de encuestas realizadas a principios de la misma. De hecho, Crespi (1988) encontró que la proximidad de las elecciones fue uno de los factores más importantes para determinar la precisión de la encuesta. (Desart y Holbrook, 2003, p. 432)

Por su parte, Jungherr, J'Urgens y Schoen (2011) critican la labor de Tumasjan et al., (2010). Ellos argumentan que la elección de los partidos incluidos

¹⁹⁷ Afirmación realizada por la investigadora Mariluz Congosto obtenida sobre su estudio de la identificación de voto mediante análisis de Twitter (proyecto de investigación en curso en la actualidad, mayo de 2012).

en la evaluación no fue la correcta y demuestran que la introducción de un séptimo partido, el Partido Pirata, habría tenido un efecto negativo sobre la exactitud de las predicciones. Además, cuestionan el período de tiempo que fue utilizado por los investigadores para la toma de datos (captura de *tweets*), demostrando que la inclusión de los *tweets* en la semana antes de las elecciones también afectaría negativamente a la predicción que aquéllos hallaron.

Los hallazgos de estos autores, aunque con base en distintos contenidos (encuestas y Twitter, respectivamente) se contradicen. Para los primeros, la cercanía de la encuesta a la celebración de los comicios es un factor positivo, mientras que para los segundos, esta aproximación lograría un efecto contrario.

Debido a estos hallazgos se pone de manifiesto la relevancia de una nueva línea de investigación que analizara los datos conseguidos en Twitter en diferentes momentos a los previamente tomados aquí en consideración. La comparación de resultados Twitter en una misma elección, obtenidos a través de diferentes franjas de tiempo en la toma de datos podría ayudar a desvelar cuál es el momento más indicado para el análisis, lo cual es de vital relevancia para determinar el correcto uso de este tipo de análisis en Redes Sociales.

4.3. Metodología para la captura de datos

Cada elección es un objeto autónomo e independiente del resto. En cada caso concreto, aunque en todos ellos se ha utilizado como herramienta los APIs de Twitter¹⁹⁸, la metodología para la captura de los datos difiere en varios aspectos:

- Las fechas de inicio y fin del periodo establecido para la captura, dependientes de la fecha de la jornada electoral.
- Los términos de búsqueda, establecidos sobre los nombres de los perfiles Twitter de candidatos y partidos, y los *hashtag* de sus respectivas campañas electorales.

Se hace necesario constatar la dificultad que entraña este tipo de análisis, derivada de la particularidad de que los datos que no se capturan en el momento en que la actividad o conversación está teniendo lugar, son datos perdidos y su análisis, por tanto, irrealizable. Para analizar un hecho determinado en Twitter, es preciso que la vigilancia y descarga de la información circulante se lleve a cabo en tiempo real, cuando está sucediendo¹⁹⁹.

A continuación detallamos la metodología seguida en cada proceso²⁰⁰.

¹⁹⁸ Se han utilizado, según criterio de la investigadora Mariluz Congosto (2010-2011), las tres APIs de Tittwer para desarrolladores: Search API, Rest API y Streaming API. La primera, Search API, es el API diseñado para buscar a un usuario para consultar el contenido de Twitter. como la búsqueda de una serie de tweets que contienen palabras clave específicas, la de los tweets que hacen referencia a un usuario específico, o la de tweets publicados por un usuario en particular. El REST API, que puede ser utilizada como herramienta de escritorio, permite acceder a información detallada de un usuario en particular, como su línea de tiempo, sus actualizaciones en Twitter y los detalles de su perfil. Por último, el Streaming API, es una herramienta de captura pensada para desarrolladores con necesidades intensivas de datos, para aquellos casos en los que el caudal de tweets publicados es tan grande que el Seach API no es capaz de capturarlos en su totalidad, siendo la herramienta Twitter más recomendada para proyectos de investigación. Más información en <https://dev.twitter.com/start>

¹⁹⁹ La empresa Twitter almacena todos los datos en sus servidores, siendo posible intentar la compra de los mismos, según condiciones que ella misma marque.

²⁰⁰ En este partado se computan los volúmenes globales de las capturas realizadas en cada proceso electoral. En el capítulo de análisis se menciona el volumen de *tweets* referido únicamente al periodo de campaña electoral.

4.3.1. Elecciones Generales 2011

Periodo de captura: desde el 4 de noviembre de 2011 hasta el 19 de noviembre de 2011.

Volumen de datos: se han obtenido 1.592.872 *tweets*.

- **Dataset Total:** es la suma de los *datasets* Lemas y Candidatos y Debate, una vez eliminados los *tweets* repetidos.
 - **Dataset Lemas:** capturó todos los *tweets* que contenían algunas de los siguientes términos: #20n, #movimientopopular, #empiezaelcambio, #rubalcabasi, #equova, #nolesvotes, #democracialrealya, #spanishrevolution, #15m, #15°, #psoe, #pp, #ppopular, #iu, #iunida, #upyd, rajoy, rubalcaba, cayo lara, rosa díez.
 - **Dataset candidatos:** obtuvo los *tweets* de los usuarios: @conRubalcaba, @marianorajoy, @cayo_lara, @psoe, @ppopular, @iunida, @upyd, @Proyectoequo, @democraciareal, @acampadasol y los *tweets* que incluían alguna mención a ellos.
 - **Dataset debate:** recogió los *tweets* que incluían las palabras: #ReiniciaElDebate, #seacaboelcirco, #ocupaeldebate, #undebatedecisivo, #debate2011, #debate, #caraacara, rubalcaba, Rajoy.

Excepciones: se recolectaron los datos por tres vías distintas. En unos casos para separar información y en otros por limitaciones del API de Twitter que no permite monitorizar a la vez a usuarios y palabras claves. La captura de *tweets* se hizo con varios “hilos” el primero para capturar los *tweets* con lemas de la campaña, el segundo para candidatos y el tercero para el debate²⁰¹. En todas las capturas se

²⁰¹ Se utilizó en todos los casos el **Streaming API de Twitter**, que es el que permite capturar más caudal de *tweets*, sin embargo hay que tener en cuenta que en el *dataset* del debate se alcanzó la velocidad máxima de 50 *tweets* por segundo (TPS) por lo que la muestra no es totalmente completa.

recogió el identificativo de cada *tweet* para que al fusionarlos no se vieran *tweets* repetidos.

4.3.2. Elecciones Autonómicas a la Comunidad de Madrid 2011

La captura de datos en Twitter se ha llevado a cabo de la siguiente manera:

Periodo de captura: desde el 6 mayo de 2011 hasta el 21 de mayo de 2011.

Volumen de datos: se han obtenido 134. 193 *tweets*.

- **Dataset 1:** es el resultado de la fusión de los datasets 1 y 2, una vez eliminados los *tweets* repetidos.
 - **Dataset 1:** contiene los *tweets* globales del resultado de la búsqueda de las menciones a los candidatos y a los partidos que se presentaban a las elecciones de la Comunidad de Madrid el 22 de mayo del 2011, junto con los a *hashtags* locales.
 - **Dataset 2:** recoge los tweets de **Madrid** que referencian a partidos o a *hashtags* globales durante la monitorización de la campaña.

Términos de búsqueda: @contigotomas, @psmadrid, @jsmadrid, @ppmadrid, @esperanzaguirre, @gregorio_gordo, @iucm, @Jovenesiucm, @develascorami, @victordomingo, #elcambioposible, #contigotomas, #psmadrid, #jsmadrid, #ppmadrid, #esperanzaguirre, #iucm, #centradosenti, #sobranmotivos, #gregorio_gordo, PSRM, PSOE, PP, IU, UPyD_Madrid, @upyd_Madrid, #upyd_Madrid, #mareamagenta, #votoutil, @democraciareal, @nolesvotes, #nolesvotes, #democraciareal

4.3.3. Elecciones Autonómicas a la Región de Murcia 2011

Periodo de captura: desde el 6 de mayo de 2011 hasta el 21 de mayo de 2011.

Volumen de datos: se han obtenido 3.303 *tweets*.

- **Dataset total:** es el resultado de la fusión de los datasets 1 y 2, una vez eliminados los *tweets* repetidos.
 - **Dataset 1:** contiene los *tweets* globales del resultado de la búsqueda de las menciones a los candidatos y a los partidos que se presentaban a las elecciones de la Asamblea de la Región de Murcia el 22 de mayo del 2011, junto con los *hashtags* locales.
 - **Dataset 2:** Hashtags globales, sólo datos referidos a la localización Murcia

Términos de búsqueda: “Rafael Sanchez”, @rasadi2, @UPyDMurcia ,”Begoña García Retegui”, @garciaretegui, PSOEMURCIA, @psrmpsoe , @jsrmurcia , “Ramón Luis Valcárcel” , @PPMurcia , @popularesRM , Pujante , @IUVRM , @japujante, #PSRM, #dalelavuelta, #garciaretegui, #PSOEMURCIA, #psrmpsoe, #jsrmurcia, #rasadi2, #PPMurcia, #IUVRM, #japujante”, #votoutil, #mareamagenta, #transparencia, #mueveteiuverdes, @rasadi2, @garciaretegui, @japujante, @UPyDMurcia, @PSOEMURCIA, @psrmpsoe, @jsrmurcia, PPMurcia, @popularesRM, @IUVRM.

4.3.4. Elecciones al Parlamento de Cataluña 2010

La captura de datos en Twitter se ha llevado a cabo de la siguiente manera:

Periodo de captura: desde el 12 de noviembre de 2010 hasta el 27 de noviembre de 2010.

Volumen de datos: se han obtenido 72.485 *tweets*.

Términos de búsqueda²⁰²: 28N, presidentmontilla, eleccions, eleccions10, eleccions2010, CiU, PPCatalunya, esquerra, ERC,ICV, icveuia, PSC, socialistes, cativistes, debatoliva, eleccionscat, solidaritatcatalana, programaPSC, socialistes_cat, solucions, populars, Artur Mas, Montilla, Puigcercós, puigcercos2010, Joan Herrera, Sánchez Camacho, Sánchez-Camacho, Albert Rivera, albert_rivera, rebelate, Maspresident.

Excepciones: dado que PSC y CIU son términos usados en otros países de los *tweets* iniciales recogidos se descartaron aquellos que no estaban en catalán o en castellano.

4.4. Acerca de la validez de la muestra

La selección de la muestra es uno de los factores a tener en cuenta. Las muestras tienen que ser representativas y el muestreo obedecer a la aleatoriedad, lo que garantiza la posibilidad de extrapolar mediante la inferencia estadística:

- la máxima aleatoriedad de la selección de los individuos que la componen, esto es, que todos los individuos de una elección dada tengan las mismas probabilidades de ser elegidos para la encuesta.
- si el conjunto de una población, la población objeto de estudio (toda la población electoral: españoles mayores de 18 años en una circunscripción dada) se encuentra suficientemente representada.

Respecto a la primera, la muestra que obtenemos en los datos recogidos en Twitter cumple en parte el requisito de la aleatoriedad ya que no sólo no se produce selección de usuarios sino que se “encuesta” toda la población de dicha Red que ha participado con sus comentarios en la campaña electoral.

²⁰² En el caso de las Elecciones al Parlamento de Cataluña 2010, y debido a que fue la primera vez que monitorizábamos procesos electorales en Twitter, no se capturaron los perfiles de los candidatos.

Para analizar si el marco muestral de Twitter es representativo de la población objeto de estudio (población mayor de 18 años en una circunscripción dada), se ha establecido una comparación²⁰³ entre el número de individuos de una población determinada (datos del padrón del INE, Instituto Nacional de Estadística, 2011) y los usuarios de Twitter en esa misma población (Datos de Elogia-IAB, 2011, y Adigital, 2010). En los casos en los que se ha podido por disponer de la información necesaria, se han establecido comparaciones por sexo y franja de edad, con segmentos habitualmente empleados para la configuración de muestras en encuestas demoscópicas de carácter electoral²⁰⁴.

Téngase en cuenta la limitación y alcance de esta comparación. La población mayor de 55 años no se encuentra representada en esta Red. Se supone que el índice de participación en Redes Sociales en esa franja de edad es prácticamente irrelevante. En los estudios sobre uso de Twitter en España, las encuestas no suelen incluir individuos que sobrepasen esta edad, debido seguramente a que el uso de esta herramienta social es muy escaso en esta franja de la población.

En todas las circunscripciones electorales analizadas (los cuatro procesos electorales objeto de análisis) el método que se ha empleado es el siguiente:

- Cálculo de la población con edades comprendidas entre los 18 y los 55 años. Se han sumado todos los individuos.
- Cálculo de la población usuaria de Twitter con edades comprendidas entre los 18 y los 55 años, obtenida sobre el porcentaje de internautas - se considera internautas a los usuarios frecuentes de Internet (al menos 1 vez por semana)- que son usuarios de esta red en dicha circunscripción.

²⁰³ Hacemos alusión aquí al análisis cuantitativo. El análisis descriptivo se incluyó en el epígrafe del Marco Teórico destinado al “Grado de Penetración de las Redes Sociales en España”.

²⁰⁴ Sólo se han encontrado referencias por sexo y edad en el uso de Twitter a nivel nacional.

- Obtención del valor en porcentaje de la relación “Población usuaria de Twitter/Población INE”, siendo mejor el dato obtenido cuanto más cerca del 100% se encuentre el valor que se obtenga.

Los resultados, por circunscripciones, son:

Tabla 14. Comparación del marco muestral en las Elecciones Generales 2011

Elecciones Generales 2011			
	Población INE	Usuarios Twitter	% población en Twitter
Global	26177427	7888000	30,13
Hombres	13313133	4811680	36,14
Mujeres	12864294	3076320	23,91
De 16 a 18 años	1361163	709920	52,16
De 19 a 25 años	3675916	2366400	64,38
De 26 a 35 años	5894524	2603040	44,16
Más de 35 años	14444521	2287520	15,84

- 23,2 millones de españoles conectados a Internet (INE 2011)
- 34% de internautas utilizan Twitter (Elogia e IAB, 2011)
- Por sexo, 61% hombre y 39% mujeres (The Cocktail Analysis, 2011)
- Por edad,
 - De 16 a 18 años: 9%,
 - De 19 a 25 años: 30%
 - De 26 a 35 años: 33%
 - Más de 35 años: 29%

Fuente: elaboración propia

Como se ha apuntado anteriormente, para el resto de entornos electorales no se dispone de datos segmentados por franja de edad o sexo, por lo que las tablas que a continuación se muestran deben tomarse únicamente como referencia de la relación del porcentaje de población con derecho a voto que participa en la Red Social Twitter.

Tabla 15. Comparación del marco muestral en las Elecciones Autonómicas a la Comunidad de Madrid 2011

Elecciones Comunidad de Madrid 2011		
Población INE	Usuarios Twitter	% población en Twitter
3538353	980267	27,70

- 70,8% de la población está conectada a Internet (INE 2011).
- 39,13% usan Twitter (adigital, 2010)

Fuente: elaboración propia

Tabla 16. Comparación del marco muestral en las Elecciones Autonómicas a la Región de Murcia 2011

Elecciones Región de Murcia 2011		
Población INE	Usuarios Twitter	% población en Twitter
786425	6443	0,82

- 56,5% de la población está conectada a Internet (INE 2011).
- 1,45% usan Twitter (adigital, 2010)

Fuente: elaboración propia

Tabla 17. Comparación del marco muestral en las Elecciones al Parlamento de Cataluña 2010

Elecciones Parlamento de Cataluña 2010		
Población INE	Usuarios Twitter	% población en Twitter
3964042	780423	19,69

- 71,8% de la población está conectada a Internet (INE 2011).
- 27,42% usan Twitter (adigital, 2010)

Fuente: elaboración propia

Tal y como se desprende de los resultados obtenidos, las relaciones de la comparación de los marcos muestrales no revelan buenas posiciones puesto que todas ellas se encuentran alejadas del valor deseado, el 100%, y siendo la muestra tomada a nivel nacional la que obtiene el mejor valor y la de la Región de Murcia muy por debajo de un valor mínimamente aceptable.

En el capítulo dedicado a los Resultados del Análisis se intentará confirmar o desdecir si también en Twiter el tamaño de la muestra es un factor relevante para la predicción, tal y como concluyeron Desart y Holbrook (2003).

Es preciso indicar que en el caso de investigación que nos ocupa, bien sería posible que en la eficiencia de la predicción electoral mediante el análisis de la actividad Twitter, el número de individuos en una circunscripción dada que utiliza o se define como usuario de Twitter podría no ser el factor determinante, y que ésta podría depender en cierta medida del número de usuarios únicos que participaron en una determinada campaña, y del volumen de actividad en dicha Red (medida en número de *tweets*).

Para concluir este epígrafe, hacemos referencia a las limitaciones de este análisis. Además del hecho de que la totalidad de la población con derecho a voto no es usuaria de Twitter, la muestra finalmente “encuestada” (población con derecho a voto que ha participado en la campaña electoral en Twitter) no viene segmentada ni por edad ni por sexo (no todos los usuarios de esta red indican en su perfil su sexo y su edad).

Capítulo **4**

Resultados de la Investigación

5. Capítulo IV: Resultados de la investigación

5.1. Introducción

El presente capítulo muestra los resultados finales de la investigación obtenidos para cada una de las elecciones monitorizadas: Elecciones Generales 2011, Elecciones Autonómicas a la Comunidad de Madrid 2011, Elecciones Autonómicas a la Región de Murcia 2011 y Elecciones al Parlamento de Cataluña 2010.

Para todas ellas, tal y como se referenció en el capítulo dedicado a la Metodología, se han sumado de forma independiente todas y cada una de las actividades twitter, tomadas como variables, que una determinada opción política ha recibido en esta Red durante la campaña electoral, señalando las fechas de captura en cada caso. Se indica, por tanto, para cada una de las variables establecidas los *strings* o cadenas de texto que se han tomado como referencia para realizar el cómputo global, esto es, se han sumado aquellos *tweets* en los que dichas cadenas de texto aparecían.

Para la primera y segunda línea de investigación, lo análisis que se han llevado a cabo en cada una de las elecciones y cuyos resultados finales se incorporan en sendas tablas, son los siguientes:

- Resultado Twitter
- Resultado Twitter Sólo Menciones
- Resultado Twitter Candidato
- Resultado Twitter Candidato Sólo Menciones
- Resultado Twitter Partido
- Resultado Twitter Partido Sólo Menciones

Dado que en Twitter es posible referenciar tanto a candidatos como a partidos, se han efectuado diferentes cálculos, bien sea teniendo en cuenta el total de la actividad o la actividad hallada para candidatos o partidos separadamente.

Además, con el objeto de seguir la metodología empleada por Tumasjan et al., (2010), en cada uno de los casos, se ha calculado el resultado teniendo en cuenta las menciones de forma única y evitando el resto de actividad Twitter, como *Hashtag*, *RT*, *Reply* o *tweets* publicados por sus protagonistas.

Puesto que el objetivo final de esta investigación no es tanto realizar una proyección de voto sino establecer una comparación entre los resultados electorales y las tendencias electorales apreciadas en la actividad de Twitter, el porcentaje²⁰⁵ sobre el que se estima ese posible escenario de resultado electoral no es el 100% sino la suma del porcentaje de voto que obtienen los partidos políticos monitorizados. De este modo, se trata de evitar incurrir en un sesgo de selección.

Si bien hemos señalado que no se pretende tanto una proyección de voto como una comparación de resultados, para completar el alcance y conclusiones de esta investigación se ha establecido una comparación entre los resultados que la vigilancia de esta Red y los que arrojan las encuestas de opinión tradicionales en los cuatro escenarios electorales analizados. De este modo, se presentan en sendas tablas, las diferencias que ambos métodos han obtenido en relación con los resultados electorales definitivos.

Para completar el análisis, se ha llevado a cabo, una aproximación a la tercera línea de investigación, referida ésta a un cambio en el periodo de datos capturados, reduciéndose de quince a siete los días de recogida de datos, siendo estas jornadas las siete justamente anteriores a la celebración electoral. Podrá apreciarse que sólo se incluyen los resultados de dos de las cuatro elecciones monitorizadas. Los procesos electorales elegidos han sido las Elecciones Generales 2011 y las Elecciones

²⁰⁵ Dado que se ha vigilado y capturado la actividad en esta Red para los partidos que han alcanzado mayor representación parlamentaria, obviándose el resto de participantes, se ha estimado por tanto que dicha actividad no supone el 100% de la actividad política de los ciudadanos en Twitter para una campaña dada, ajustándose su valor global igualándolo a la suma del porcentaje de votos obtenidos por los partidos monitorizados.

Autonómicas a la Comunidad de Madrid y los motivos para justificar dicha selección y la no continuidad con el resto de procesos, han sido los siguientes:

- En ambos comicios las cuatro opciones políticas que obtuvieron mayor representatividad parlamentaria y que fueron monitorizadas, hicieron uso de esta Red²⁰⁶.
- Las Elecciones Generales 2011 han sido los únicos procesos electorales en los que se dan cita el conjunto de necesidades para llevar a cabo un análisis Twitter con cierto rigor. Se ha decidido incorporar las Elecciones Autonómicas a la Comunidad de Madrid, con el objetivo de establecer una comparación (si bien se entiende que los escenarios no son del todo comparables) y no dejarse llevar por las conclusiones obtenidas del análisis de uno sólo de los casos.
- Las Elecciones Autonómicas a la Región de Murcia, dado que se conoce que la opción ganadora carecía de presencia en esta Red, ha sido desestimada.
- Igualmente, las Elecciones al Parlamento de Cataluña 2010 han sido desestimadas de esta línea de investigación, ya que no se dispone de los datos Twitter referidos a ninguno de los candidatos electos.

Esta tercera línea de análisis debe tomarse simplemente como una primera aproximación a la investigación comparativa de datos recogidos en diferentes fechas cercanas o lejanas a las jornadas electorales, así como un acercamiento a la definición acertada de los días en los que estos datos pudieran resultar más eficaces.

Por otro lado, el análisis de datos en diferentes periodos bien podría resultar de utilidad para profesionales y académicos dedicados al estudio y seguimiento de campañas electorales, al mostrarnos como éstas se desarrollan en el tiempo, pudiendo

²⁰⁶ Con la excepción de la candidata Rosa Díez, del partido UPYD, por las razones que ya se han mencionado en este trabajo.

observar de primera mano y con un coste relativamente sencillo, qué posiciones van alcanzando cada uno de los candidatos en la contienda.

5.2. Análisis primera y segunda línea de investigación

5.2.1. Elecciones Generales 2011

5.2.1.1. Datos de Partida Elecciones Generales 2011

La Tabla 18 recoge la suma del número diario de acciones Twitter (variables) por opción política desde el día 4 del 11 de 2011 hasta el día 19 del mismo mes. Las opciones políticas que se han monitorizado son, como se muestra en la Tabla 19: PSOE (Partido Socialista Obrero Español), PP (Partido Popular), IU-LV (coalición Izquierda Unida-Los Verdes) y UPYD (Unión Progreso y Democracia). Los valores que se recogen por cada una de ellas son “absoluto” (suma de los valores obtenidos diariamente) y “ponderado” (resultado de aplicar la ponderación indicada al valor absoluto obtenido).

El volumen de actividad o número de *tweets* es de 1.592.872.

Los términos monitorizados para cada opción política fueron los siguientes:

Tabla 18. Relación de términos monitorizados en las Elecciones Generales 2011

	PSOE	PP	IU-LV	UPYD
Candidato	@conRubalcaba	@marianorajoy	@cayo_lara	(eliminado) ²⁰⁷
Partido	@psoe	@ppopular	@iunida	@upyd
Menciones a	@conRubalcaba	@marianorajoy	@cayo_lara	#rosadiez

²⁰⁷ La candidata de UPYD, Rosa Díez abandonó su cuenta en Twitter en marzo de 2009, pocos días después de haber conseguido un escaño en las elecciones del País Vasco, a raíz de las críticas que suscitó que estuviera en dos sitios a la vez: “La virtual Rosa Díez estaba respondiendo a las felicitaciones por sus resultados electorales mientras que, en el mismo instante, la de carne y hueso respondía a las preguntas de la entrevista, en directo, que la periodista Pepa Bueno le hacía en TVE.” <http://www.publico.es/espana/206883/rosa-diez-esta-offline>. (Consultado el 18/02/2011)

candidato	#conrubalcaba #rubalcaba rubalcaba	#rajoy #marianorajoy rajoy	#cayo_lara #cayolara cayo_lara	#rosa_diez rosadiez
Menciones a partido	@psoe #psoe psoe	@ppopular #ppopular popular pp	@iunida #iunida iunida iu	@upyd #upyd upyd
Hashtags de campaña	#rubalcabasi	#empiezaelcambio #movimientopopular	#tueliges	#cadavotovale

Fuente: elaboración propia

Tabla 19. Datos globales Twitter en las Elecciones Generales 2011

Twitter en las Elecciones Generales 2011									
Ponderación	Variables	PSOE		PP		IU-LV		UPYD	
		Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado
1	Followers candidato	26391	26391,00	35113	35113,00	3513	3513,00	0	0,00
1	Followers partido	4517	4517,00	9067	9067,00	1999	1999,00	6248	6248,00
1	Tweets publicados por candidato	1418	1418,00	833	833,00	65	65,00	0	0,00
1	Tweets publicado por partido	191	191,00	333	333,00	1532	1532,00	1830	1830,00
3	Menciones a candidato (en @, en # y en texto)	627648	1882944,00	883901	2651703,00	16409	49227,00	19434	58302,00
5	Menciones a partido (en @, en # y en texto)	68756	343780,00	135277	676385,00	33858	169290,00	53047	265235,00
9	Menciones a hashtag de campaña	908	8172,00	1044	9396,00	1414	12726,00	6800	61200,00
4	RT a candidatos	17947	71788,00	13452	53808,00	1576	6304,00	0	0,00
6	RT a partidos	730	4380,00	3730	22380,00	7285	43710,00	10364	62184,00
2	Reply a candidatos	25142	50284,00	24120	48240,00	1176	2352,00	0	0,00
4	Reply a partidos	1057	4228,00	2106	8424,00	1616	6464,00	4011	16044,00
SUMA		774705	2398093,00	1108976	3515682,00	70443	297182,00	101734	471043

Fuente: elaboración propia

5.2.1.2. Resultados de la Investigación Elecciones Generales 2011

En este caso, como puede verse en la siguiente tabla (Tabla 20), y aunque los resultados absolutos son aceptables, encontrándose una desviación máxima del 4%, el valor que se obtiene tras aplicar la ponderación a los mismos mejora sustancialmente en todos los casos salvo para la opción UPYD. Comparando con los resultados electorales, que para estas elecciones se han tomado de la página Web del Ministerio del Interior, http://www.generales2011.mir.es/99CG/DCG99999TO_L1.htm (Consultada el 21 del 11 de 2011), la mayor de la desviaciones (IU-LV) apenas supera el 3%,

porcentaje que se encuentra en el rango de error permitido o asumido por las encuestas tradicionales.

Tabla 20. Resultado Twitter en las Elecciones Generales 2011²⁰⁸

Resultado Twitter en las Elecciones Generales 2011							
	Suma Ponderada	%Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSOE	2398093	30,49	1,76	28,73	774705	32,02	3,29
PP	3515682	44,70	0,08	44,62	1108976	45,83	1,21
IU-LV	297182	3,78	-3,14	6,92	70443	2,91	-4,01
UPYD	471043	5,99	1,30	4,69	101734	4,20	-0,49
SUMA	6682000	84,96		84,96	2055858	84,96	

Fuente: elaboración propia

Con el objeto de seguir fielmente la metodología empleados por Tumasjan et al., (2010) en su investigación de la predicción de Twitter en las elecciones alemanas, se ha repetido el cálculo teniendo en cuenta las variables V5 y V6, que representan las menciones a candidatos y a partidos, respectivamente. Como puede observarse en la Tabla 22, el resultado obtenido es similar que en el caso de contar con el conjunto de variables aunque con una desviación mayor.

Tabla 21. Datos globales Twitter en las Elecciones Generales 2011 – Sólo menciones

Twitter en las Elecciones Generales 2011 - Sólo menciones									
Ponderación	Variables	PSOE	Ponderado	PP	Ponderado	IU	Ponderado	UPYD	Ponderado
2	Menciones a candidato (en @, en # y en texto)	627648	1255296,00	883901	1767802,00	16409	32818,00	19434	38868,00
4	Menciones a partido (en @, en # y en texto)	68756	275024,00	135277	541108,00	33858	135432,00	53047	212188,00
	SUMA	696404	1530320,00	1019178	2308910,00	50267	168250,00	72481	251056

Fuente: elaboración propia

²⁰⁸ Salvo las columnas “Suma Ponderada” y “Suma Twitter”, todos los datos vienen expresados en porcentajes. Idem para resto de tablas y procesos monitorizados.

Tabla 22. Resultado Twitter en las Elecciones Generales 2011 – Sólo menciones

Resultado Twitter en las Elecciones Generales 2011 - Sólo Menciones							
	Suma Ponderada	%Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Twitter	Diferencia Sin Ponderación
PSOE	1530320,00	30,53	1,80	28,73	696404	32,18	3,45
PP	2308910,00	46,06	1,44	44,62	1019178	47,10	2,48
IU-LV	168250,00	3,36	-3,56	6,92	50267	2,32	-4,60
UPYD	251056,00	5,01	0,32	4,69	72481	3,35	-1,34
SUMA	4258536,00	84,96		84,96	1838330	84,96	

Fuente: elaboración propia

Profundizando en el seguimiento de esta metodología en la que se monitorizaban únicamente menciones por partido, replicamos el cálculo inicial, en el que se toman en cuenta todas las variables²⁰⁹, pero aplicándolas únicamente al caso de los candidatos. Como puede verse en la Tabla 23, los resultados, con o sin ponderación, están más alejados de los resultados electorales.

Tabla 23. Resultado Twitter en las Elecciones Generales 2011 – Candidato

Resultado Twitter en las Elecciones Generales 2011 - Candidato							
	Suma Ponderada	%Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Twitter	Diferencia Sin Ponderación
PSOE	2032825	34,95	6,22	28,73	698546	34,95	6,22
PP	2789697	47,96	3,34	44,62	957419	47,90	3,28
IU-LV	61461	1,06	-5,86	6,92	22739	1,14	-5,78
UPYD	58302	1,00	-3,69	4,69	19434	0,97	-3,72
SUMA	4942285	84,96		84,96	1698138	84,96	

Fuente: elaboración propia

Aplicando la opción de “sólo menciones” para el perfil “candidato”²¹⁰, los resultados obtenidos, al igual que ocurría en el cálculo inicial que toma en cuenta todas las variables, son similares a aquellos que tienen en cuenta todas las variables para las menciones de los candidatos, aunque con una pequeña desviación que es mayor cuando se suman todas las del conjunto de resultados de las opciones políticas.

²⁰⁹ No se incluyen menciones a hashtags de campaña.

²¹⁰ En este caso, al ser una única la variable analizada, no se calculan los valores con ponderación al resultar inútiles (matemáticamente, arrojan idéntico resultado que “sin ponderación”)

Tabla 24. Resultado Twitter en las Elecciones Generales 2011 – Sólo menciones - Candidato

Resultado Twitter las Elecciones Generales 2011 - Sólo Menciones - Candidato				
	Resultado Electoral	Suma Twitter	% Twitter	Diferencia Sin Ponderación
PSOE	28,73	627648	34,46	5,73
PP	44,62	883901	48,53	3,91
IU-LV	6,92	16409	0,90	-6,02
UPYD	4,69	19434	1,07	-3,62
SUMA	84,96	1547392	84,96	

Fuente: elaboración propia

A continuación, repetimos los cálculos referenciados para la opción única de “partidos”. Tal y como puede observarse en las tablas referidas a “candidatos”, los resultados que se muestran en las Tablas 25 y 26 son peores que los obtenidos en el cálculo inicial que toma en cuenta tanto a candidatos como a partidos, e incluso que los obtenidos para la opción única “candidatos”.

Tabla 25. Resultado Twitter en las Elecciones Generales 2011 – Partido

Resultado Twitter en las Elecciones Generales 2011 - Partido							
	Suma Ponderada	%Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Twitter	Diferencia Sin Ponderación
PSOE	357096	18,41	10,32	28,73	75251	18,40	10,33
PP	716589	36,94	7,68	44,62	150513	36,79	7,83
IU-LV	222995	11,49	-4,57	6,92	46290	11,32	-4,40
UPYD	351541	18,12	-13,43	4,69	75500	18,46	-13,77
SUMA	1648221	84,96		84,96	347554	84,96	

Fuente: elaboración propia

Tomando únicamente los valores referidos a la opción “partido”, se observa que las desviaciones son todavía mucho mayores que en el caso en el que se tomaba sólo en cuenta la opción “candidato”. Este resultado contradice, en parte, los resultados obtenidos por Tumasjam et al. (2010), en las que se tenía en cuenta a los partidos, y puede aseverar las conclusiones del estudio de Gayo-Avello (2011) en el que indica que los métodos no pueden replicarse de una a otra elección de forma fácil y directa dados los factores que diferencian unos contextos electorales de otros.

En nuestro caso y como puede comprobarse en la siguiente tabla (Tabla 26), tampoco en este caso la toma en cuenta única de “menciones a partido”, mejora los resultados, sino todo lo contrario, encontrándose desviaciones considerables.

Tabla 26. Resultado Twitter en las Elecciones Generales 2011 – Sólo menciones - Partido

Resultado Twitter las Elecciones Generales 2011 - Sólo Menciones - Partido				
	Resultado Electoral	Suma Twitter	% Twitter	Diferencia Sin Ponderación
PSOE	28,73	68756	20,08	8,65
PP	44,62	135277	39,50	5,12
IU-LV	6,92	33858	9,89	-2,97
UPYD	4,69	53047	15,49	-10,80
SUMA	84,96	290938	84,96	

Fuente: elaboración propia

5.2.1.3. Conclusiones Análisis Twitter en las Elecciones Generales 2011

Monitorizar sólo partidos o sólo candidatos no ofrece buenos resultados. Éstos mejoran indudablemente cuando se toman en cuenta ambas facciones. Al analizarlas por separado, la opción “partido” arroja mayores desviaciones que la opción “candidato”.

Los partidos mayoritarios, PSOE y PP, alcanzan resultados muy cercanos a los resultados electorales. Las mayores desviaciones se encuentran en los partidos con menor representación parlamentaria, apareciendo la opción IU-LV infradimensionada y la candidatura UPYD, por el contrario, supradimensionada.

Es también objeto de las conclusiones de este trabajo de investigación observar si los patrones observados en estas elecciones se repiten en el resto de procesos monitorizados.

5.2.1.4. Encuestas Electorales versus Análisis Twitter en las Elecciones Generales 2011

Con el fin de establecer una comparación entre la posible capacidad de eficacia de este método en la predicción de resultados electorales y la eficacia demostrada, en mayor o menor medida, por las encuestas o métodos tradicionales, se han recogido los resultados de las encuestas preelectorales publicadas por los medios de comunicación y la encuesta preelectoral sondeada por el CIS²¹¹.

Los resultados de las encuestas se han obtenido en las siguientes direcciones Web²¹²:

CIS	http://www.mpr.es/uploads/media/pdf/0/preelectoral-elecciones-generales-2011_1320405867.pdf
Metroscopia para el diario El País	http://www.elpais.com/graficos/espana/Intencion/voto/elpepunac/20111112elpepunac_1/Ges/
SIGMADOS para el diario El Mundo	http://elmundo.orbyt.es/2011/11/13/elmundo_en_orbyt/1321139849.html
NC Report para el diario La Razón	http://www.larazon.es/noticia/9900-analisis-monopartidismo-electoral
DYM para el diario ABC	http://www.abc.es/20111113/espana/abcp-afianza-mayoria-absoluta-supera-20111113.html

Tanto los resultados de la encuestas como los obtenidos en el análisis de Twitter, se expresan en “% de intención de voto”. Como muestra de resultados Twitter, se han tomado los obtenidos después de aplicar la ponderación.

²¹¹ Centro de Investigaciones Sociológicas. España.

²¹² Consultadas el 05 de abril de 2012.

En la primera tabla (Tabla 27) se indican los valores obtenidos por cada una de las encuestas para cada una de las cuatro opciones políticas estudiadas. En la siguiente tabla (Tabla 28), se muestran las diferencias absolutas para cada una de ellas, observándose que los pronósticos obtenidos mediante el seguimiento de Twitter permiten ser comparados con los que reflejan las distintas encuestas, siendo para los partidos mayoritarios de validez semejante e incluso mejor.

El MAE en Twitter es de un 1,57%, mejorando la media de desviación de las encuestas realizadas por Sigma Dos y DYM, y muy próximo al obtenido por las encuestas del CIS, Metroscopia y NC Report.

Tabla 27. Pronósticos encuestas versus resultados Twitter en las Elecciones Generales 2011²¹³

Elecciones Generales 2011 - Encuestas versus Resultado Ponderado Twitter							
Pronósticos Encuestas							
CANDIDATURAS	CIS	Metroscopia	Sigma Dos	NC Report	DYM	Twitter	Resultados
PSOE	29,91	30,9	29,8	31	34,2	30,49	28,73
PP	46,6	45,4	47,6	45,9	46,5	44,70	44,62
IU-LV	6,17	8,8	5,7	7,2		3,78	6,92
UPYD	2,91	4,2	3	2,8		5,99	4,69
Periódicos		El País	El Mundo	La Razón	ABC		

Fuente: elaboración propia

²¹³ Todas las cifras expresan porcentajes. Los valores en este caso, con el fin de facilitar la comparación entre cantidades de las desviaciones, se han tomado en su forma “absoluta” (no diferencia entre positivo o negativo), siendo posible adquirir este dato con sólo restar de la cifra de los resultados su pareja en las difundidas por las encuestas.

Tabla 28. MAE Pronósticos encuestas versus resultados Twitter en las Elecciones Generales 2011²¹⁴

Elecciones Generales 2011 - Encuestas versus Resultado Ponderado Twitter							
Diferencia Asoluta							
CANDIDATURAS	CIS	Metroscopia	Sigma Dos	NC Report	DYM	Twitter	Resultados
PSOE	1,18	2,17	1,07	2,27	5,47	1,76	28,73
PP	1,98	0,78	2,98	1,28	1,88	0,08	44,62
IU-LV	0,75	1,88	1,22	0,28		3,14	6,92
UPYD	1,78	0,49	1,69	1,89		1,30	4,69
% MAE	1,42	1,33	1,74	1,43	3,68	1,57	

Fuente: elaboración propia

5.2.2. Elecciones Autonómicas a la Comunidad de Madrid 2011

5.2.2.1. Datos de Partida Elecciones Autonómicas a la Comunidad de Madrid 2011

La Tabla 30 recoge la suma del número diario de acciones Twitter (variables) por opción política desde el día 6 del 5 de 2011 hasta el día 21 del mismo mes. Las opciones políticas que se han monitorizado y que se muestran en la Tabla 29 son: PSM-PSOE (Partido Socialista de Madrid), PP Madrid (Partido Popular Madrid), IU-LV (coalición Izquierda Unida-Los Verdes) y UPYD Madrid (Unión Progreso y Democracia Madrid). Los valores que se recogen por cada una de ellas son “absoluto” (suma de los valores obtenidos diariamente) y “ponderado” (resultado de aplicar la ponderación indicada al valor absoluto obtenido).

El volumen de actividad o número de *tweets* es de 134.193.

Los términos monitorizados para cada opción política fueron los siguientes:

²¹⁴ Todas las cifras expresan porcentajes.

Tabla 29. Relación de términos monitorizados en las Elecciones Autonómicas a la Comunidad de Madrid 2011

	PSM- PSOE	PP Madrid	IU-LV	UPYD Madrid
Candidato	@contigotomas	@esperanzaguirre	@gregorio_gordo	@develascorami
Partido	@psmadrid	@ppmadrid	@iucm	@upyd_madrid
Menciones a candidato	@contigotomas #contigotomas tomás gómez	@esperanzaguirre #esperanzaguirre esperanza aguirre	@gregorio_gordo #gregorio_gordo gregorio gordo	@develascorami #develascorami luís de velasco
Menciones a partido	@psmadrid #psmadrid psmadrid #psm psm	@ppmadrid #ppmadrid ppmadrid	@iucm #iucm iucm	@upyd_madrid #upyd_madrid upyd_madrid
Hashtags de campaña	#elcambioposible	#centradosenti	#sobranmotivos	#mareamagenta ²¹⁵

Fuente: elaboración propia

Tabla 30. Datos globales Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011

Twitter en las Elecciones Autonómicas a la CAM 2011									
Ponderación	Variables	PSM		PP Madrid		IU-LV		UPYD Madrid	
		Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado
1	Followers candidato	595	595,00	5935	5935,00	305	305,00	205	205,00
1	Followers partido	177	177,00	958	958,00	69	69,00	136	136,00
1	Tweets publicados por candidato	1477	1477,00	192	192,00	12	12,00	216	216,00
1	Tweets publicado por partido	184	184,00	961	961,00	109	109,00	597	597,00
3	Menciones a candidato (en @, en # y en texto)	17728	53184,00	27073	81219,00	1129	3387,00	1545	4635,00
5	Menciones a partido (en @, en # y en texto)	1169	5845,00	7163	35815,00	261	1305,00	520	2600,00
9	Menciones a hashtag de campaña	1785	16065,00	3432	30888,00	290	2610,00	501	4509,00
4	RT a candidatos	888	3552,00	659	2636,00	9	36,00	14	56,00
6	RT a partidos	88	528,00	1178	7068,00	80	480,00	189	1134,00
2	Reply a candidatos	888	1776,00	2568	5136,00	46	92,00	146	292,00
4	Reply a partidos	82	328,00	751	3004,00	22	88,00	50	200,00
	SUMA	25061	83711,00	50870	173812,00	2332	8493,00	4119	14580

Fuente: elaboración propia

²¹⁵ Se ha descartado el hashtag “#votoutil” porque aunque fue muy utilizado por los simpatizantes de esta candidatura, también fue seguido por usuarios y plataformas asociados al movimiento #15M.

5.2.2.2. Resultados de la Investigación Elecciones Autonómicas a la Comunidad de Madrid 2011

En las Elecciones Autonómicas a la Comunidad de Madrid 2011, como puede verse en la Tabla 31, tanto los resultados absolutos como los ponderados sólo encuentran valores aceptables para dos de las opciones políticas monitorizadas (PSM-PSOE y UPYD Madrid), siendo las desviaciones de PP Madrid e IU-LV superiores al 6%. Dichas desviaciones, aunque, mejoran sensiblemente tras aplicar la ponderación, continúan manteniendo valores escasamente estimables. Los resultados electorales se han tomado de la página Web de Madrid.org, http://resultados2011.madrid.org/12AU/DAU12999CM_L1.htm (Consultada el 17 del 06 de 2011).

Tabla 31. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011

Resultado Twitter en las Elecciones Autonómicas a la CAM 2011							
	Suma Ponderada	%Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSM	83711	28,02	-1,75	26,27	25061	28,57	-2,30
PP Madrid	173812	58,18	-6,45	51,73	50870	58,00	-6,27
IU-LV	8493	2,84	6,77	9,61	2332	2,66	6,95
UPYD Madrid	14580	4,88	1,44	6,32	4119	4,70	1,62
SUMA	280596	93,93		93,93	82382	93,93	

Fuente: elaboración propia

Continuando con el método aplicado en el análisis de las Elecciones Generales 2011 y aunque en este caso alguno de los resultados presenta desviaciones substanciales, se procede a repetir el cálculo teniendo en cuenta las variables V5 y V6, que representan las menciones a candidatos y a partidos, respectivamente, con el fin de comprobar si los resultados mejoran tomando únicamente en cuenta las menciones. Como puede observarse en la Tabla 33, los resultados empeoran, aumentando las desviaciones en la mayoría de los casos.

Tabla 32. Datos globales Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011 – Sólo menciones

Twitter en las Elecciones Autonómicas a la CAM 2011 - Sólo Menciones									
Ponderación	Variables	PSM		PP Madrid		IU-LV		UPYD Madrid	
		Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado
3	Menciones a candidato (en @, en # y en texto)	17728	53184	27073	81219	1129	3387	1545	4635
5	Menciones a partido (en @, en # y en texto)	1169	5845	7163	35815	261	1305	520	2600
	SUMA	18897	59029	34236	117034	1390	4692	2065	7235

Fuente: elaboración propia

Tabla 33. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011-Sólo menciones

Resultado Twitter en las Elecciones Autonómicas a la CAM 2011 - Sólo Menciones							
	Suma Ponderada	% Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSM	59029	29,49	-3,22	26,27	18897	31,37	-5,10
PP Madrid	117034	58,48	-6,75	51,73	34236	56,83	-5,10
IU-LV	4692	2,34	7,27	9,61	1390	2,31	7,30
UPYD Madrid	7235	3,61	2,71	6,32	2065	3,43	2,89
SUMA	187990	93,93		93,93	56588	93,93	

Fuente: elaboración propia

Para proseguir con el método aplicado en las Elecciones Generales 2011, replicando el cálculo inicial, en el que se toman en cuenta todas las variables²¹⁶, pero aplicándolas únicamente a los candidatos. Como puede verse en la Tabla 34, los resultados, con o sin ponderación, continúan empeorando salvo para el PP de Madrid, que mejora ostensiblemente.

Tabla 34. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid - Candidato

Resultado Twitter en las Elecciones Autonómicas a la CAM 2011 - Candidato							
	Suma Ponderada	% Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSM	60584	34,50	-8,23	26,27	21576	32,88	-6,61
PP Madrid	95118	54,17	-2,44	51,73	36427	55,52	-3,79
IU-LV	3832	2,18	7,43	9,61	1501	2,29	7,32
UPYD Madrid	5404	3,08	3,24	6,32	2126	3,24	3,08
SUMA	164938	93,93		93,93	61630	93,93	

Fuente: elaboración propia

²¹⁶ No se incluyen menciones a *hashtags* de campaña.

Aplicando la opción de “sólo menciones” para el perfil “candidato”²¹⁷, los resultados se mantienen de forma similar, al igual que ocurría a aplicar la fórmula “sólo menciones” en el cálculo general, salvo para el PP de Madrid, que también en esta situación mejora notablemente.

Tabla 35. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011 – Sólo menciones - Candidato

Resultado Twitter en las Elecciones Autonómicas a la CAM 2011 - Sólo Menciones - Candidato				
	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSM	26,27	17728	35,08	-8,81
PP Madrid	51,73	27073	53,56	-1,83
IU-LV	9,61	1129	2,23	7,38
UPYD Madrid	6,32	1545	3,06	3,26
SUMA	93,93	47475	93,93	

Fuente: elaboración propia

A continuación, repetimos los cálculos referenciados para la opción única de “partidos”. Tal y como puede observarse en las tablas referidas a “candidatos”, los resultados que se muestran en las Tablas 35 y 36 son peores que los obtenidos en el cálculo inicial que toma en cuenta tanto a candidatos como a partidos, e incluso que los obtenidos para la opción única “candidatos”, mejorando sustancialmente para la opción de UPYD Madrid y ligeramente para IU-LV.

Tabla 36. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011- Partido

Resultado Twitter en las Elecciones Autonómicas a la CAM 2011 - Partido							
	Suma Ponderada	% Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSM	7062	10,77	15,50	26,27	1700	10,83	15,44
PP Madrid	47806	72,91	-21,18	51,73	11011	70,15	-18,42
IU-LV	2051	3,13	6,48	9,61	541	3,45	6,16
UPYD Madrid	4667	7,12	-0,80	6,32	1492	9,51	-3,19
SUMA	61586	93,93		93,93	14744	93,93	

Fuente: elaboración propia

²¹⁷ En este caso, al ser una única la variable analizada, no se calculan los valores con ponderación al resultar inútiles (matemáticamente, arrojan idéntico resultado que “sin ponderación”)

Tomando únicamente los valores referidos a la opción “partido”, se observa que las desviaciones son todavía mucho mayores que en el caso en el que se tomaba sólo en cuenta la opción “candidato”. Este resultado, al igual que acontecía en el análisis de Elecciones Generales 2011, no refuta los resultados obtenidos por Tumasjam et al. (2010) acerca de las elecciones alemanas.

Nuevamente en estas elecciones, al igual que en las Elecciones Generales 2011, tal y como puede observarse en Tabla 2137, tampoco la toma en cuenta única de “menciones a partido”, mejora los resultados, sino todo lo contrario, encontrándose desviaciones mucho mayores que en los casos propuestos anteriormente (general y candidato). Mejora de nuevo la opción de UPYD Madrid y se incrementan los desajustes para el resto de opciones políticas.

Tabla 37. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011 – Sólo menciones - Partido

Resultado Twitter en las Elecciones Autonómicas a la CAM 2011 - Sólo Menciones - Partido				
	Resultado Electoral	Suma Twitter	% Twitter	Diferencia Sin Ponderación
PSM	26,27	1169	12,05	14,22
PP Madrid	51,73	7163	73,83	-22,10
IU-LV	9,61	261	2,69	6,92
UPYD Madrid	6,32	520	5,36	0,96
SUMA	93,93	9113	93,93	

Fuente: elaboración propia

5.2.2.3. Conclusiones Análisis Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011

En el caso de las Elecciones Autonómicas a la Comunidad de Madrid 2011, al igual que ocurría con las Elecciones generales 2011, Monitorizar sólo partidos o sólo candidatos ofrece diferentes resultados que mejoran o empeoran en cada caso dependiendo de la opción política monitorizada. Así, cuando se toman en cuenta únicamente a “candidatos”, la candidatura PP de Madrid mejora mientras que el resto de candidaturas empeora. Las desviaciones aumentan notablemente cuando tratamos

sólo con “partidos”, caso en que sucede que mejora la candidatura de UPYD Madrid y el resto empeora en un mayor nivel que en el resto de análisis realizados (general y candidato). Anotar que la candidatura IU-LV presenta pequeñas variaciones en todos los análisis, quizás debido a su escasa presencia en esta red, lo cual, unido al hecho de que para esta opción política los resultados obtenidos hasta ahora siempre han presentado grandes desviaciones comparados con los resultados electorales, podría indicarnos que el volumen de datos es un requisito a la hora de realizar un análisis en esta red.

La opción IU-LV aparece una vez más infradimensionada y la candidatura UPYD, por el contrario, supradimensionada (aunque en esta ocasión con menores valores que los que se obtuvieron en el análisis de las Elecciones Generales 2011).

5.2.2.4. Encuestas Electorales versus Análisis Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011

Con el fin de establecer una comparación entre la posible capacidad de eficacia de este método en la predicción de resultados electorales y la eficacia demostrada, en mayor o menor medida, por las encuestas o métodos tradicionales, se han recogido los resultados de las encuestas preelectorales publicadas por los medios de comunicación y la encuesta preelectoral sondeada por el CIS²¹⁸.

Los resultados de las encuestas se han obtenido en las siguientes direcciones Web²¹⁹:

CIS ²²⁰	http://datos.cis.es/pdf/Es2871mar_A.pdf
TNS-Demoscopia para Antena 3 y	http://www.antena3.com/especiales/noticias/elecciones-autonomicas-municipales/lograria-mayoria-absoluta-comunidad-

²¹⁸ Centro de Investigaciones Sociológicas. España.

²¹⁹ Consultadas el 05 de abril de 2012

²²⁰ PREELECTORAL ELECCIONES.AUTONÓMICAS Y MUNICIPALES, 2011 (Comunidad de Madrid y Ciudad de Madrid)

Onda Cero	madrid-sevilla_2011050202145.html
Obradoiro de Socioloxia para el diario Público	http://www.publico.es/espana/373844/el-escandalo-del-caso-gurtel-no-pasa-factura-electoral-a-aguirre
GAD (Gabinete de Análisis Demoscópico) para la COPE	http://www.cope.es/espana/28-04-11--el-pp-amplia-su-mayoria-absoluta-en-madrid-240662-1

Tanto los resultados de la encuestas como los obtenidos en el análisis de Twitter, se expresan en “% de intención de voto”. Como muestra de resultados Twitter, se han tomado los obtenidos después de aplicar la ponderación.

En la primera tabla (Tabla 38) se indican los valores obtenidos por cada una de las encuestas para cada una de las cuatro opciones políticas estudiadas. En la siguiente tabla (Tabla 39), se muestran las diferencias absolutas para cada una de ellas, observándose que los pronósticos obtenidos mediante el seguimiento de Twitter permiten ser comparados las distintas encuestas únicamente para dos de las opciones políticas (PSM-PSOE y UPYD Madrid), caso en el que obtienen mejores resultados, siendo para las restantes una diferencia negativamente destacable.

El MAE de Twitter es de un 4,10%, lo que le aleja en este caso de los rangos admitidos en los estudios de opinión.

Tabla 38. Pronósticos encuestas versus resultados Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011²²¹

Elecciones Autonómicas CAM 2011 - Encuestas versus Resultado Ponderado Twitter						
Pronósticos Encuestas						
Candidaturas	CIS	TNS-Demoscopia	Obradoiro de Socioloxía	GAD	Twitter	Resultados
PSM	28,4	29,3	33,3	28,3	28,0	26,27
PP Madrid	52,1	54,5	52,3	54	58,2	51,73
IU-LV	9,1	8,6	7	10,6	2,8	9,61
UPYD Madrid	3,7				4,9	6,32
Medios		A3 y Onda Cero	Público	COPE		

Fuente: elaboración propia

Tabla 39. MAE pronósticos encuestas versus resultados Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011

Elecciones Autonómicas CAM 2011 - Encuestas versus Resultado Ponderado Twitter						
Diferencia Absoluta						
Candidaturas	CIS	TNS-Demoscopia	Obradoiro de Socioloxía	GAD	Twitter	Resultados
PSM	2,13	3,03	7,03	2,03	1,8	26,27
PP Madrid	0,37	2,77	0,57	2,27	6,5	51,73
IU-LV	0,51	1,01	2,61	0,99	6,8	9,61
UPYD Madrid	2,62				1,4	6,32
% MAE	1,41	2,27	3,40	1,76	4,10	

Fuente: elaboración propia

5.2.3. Elecciones Autonómicas a la Región de Murcia 2011

5.2.3.1. Datos de Partida Elecciones Autonómicas a la Región de Murcia 2011

En las Elecciones Autonómicas a la Región de Murcia 2011, al iniciar el seguimiento de la actividad de la campaña en Twitter, nos encontramos con varias excepciones, algunas de las cuales son de vital relevancia al obligar a obviar una de las opciones políticas. El hecho de que uno de los principales candidatos en la

²²¹ En Tablas 38 y 39, todas las cifras expresan porcentajes.

contienda, Ramón Luís Valcárcel candidato por el Partido Popular en la región, no dispusiera de perfil en esta red, y que su partido tampoco dispusiera de un perfil oficial²²², complicaba el análisis y su no monitorización reducirá la eficacia de las conclusiones del mismo ya que ni el partido político ganador ni su candidato han podido ser analizados. Por otro lado, el PSOE de esta comunidad no dispone de cuenta oficial (puesto que no se indica en su página Web <http://www.psoe-regiondemurcia.com/www/home.asp>). No obstante se tomó en cuenta el perfil # @PSOEMURCIA, perteneciente a la agrupación socialista de Murcia²²³.

La Tabla 41 recoge la suma del número diario de acciones Twitter (variables) por opción política desde el día 6 del 5 de 2011 hasta el día 21 del mismo mes. Las opciones políticas que se han monitorizado son: PSRM-PSOE (Partido Socialista de la Región de Murcia), IUV-RM (coalición Izquierda Unida-Los Verdes) y UPYD Murcia (Unión Progreso y Democracia Región de Murcia). Los valores que se recogen por cada una de ellas son “absoluto” (suma de los valores obtenidos diariamente) y “ponderado” (resultado de aplicar la ponderación indicada al valor absoluto obtenido).

El volumen de actividad o número de *tweets* es de 3.303.

Los términos monitorizados para cada opción política fueron los siguientes:

Tabla 40. Relación de términos monitorizados en las Elecciones Autonómicas a la Región de Murcia 2011

	PSRM	IUV-RM	UPYD Murcia
Candidato	@garciaretegui	@japujante	@rasadi2

²²² El perfil oficial @PPRMurcia se creó el 15 de septiembre de 2011, varios meses después de celebradas las elecciones.

²²³ Posteriormente, se han encontrado varios perfiles más (@psrmpsoe y @PSOE_RM) que o bien no habían publicado nada a finales de 2011 o crearon su cuenta meses después de celebradas las elecciones.

Partido	@PSOEMURCIA	@IUVRM	@UPyDMurcia
Menciones a candidato	@garciairetegui #garciairetegui garciairetegui	@japujante #japujante japujante	@rasadi2 #rasadi2 rasadi2
Menciones a partido	@PSOEMURCIA #PSOEMURCIA PSOEMURCIA	@IUVRM #IUVRM IUVRM	@UPyDMurcia #UPyDMurcia UPyDMurcia
Hashtags de campaña	#dalelavuelta	#mueveteiuverdes	#mareamagenta ²²⁴ #transparencia

Fuente: elaboración propia

Tabla 41. Datos globales Twitter en las Elecciones Autonómicas a la Región de Murcia 2011

Twitter en las Elecciones Autonómicas a la Región de Murcia 2011							
Ponderación	Variables	PSRM		IUVRM		UPyD Murcia	
		Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado
1	Followers candidato	52	52,00	24	24,00	56	56,00
1	Followers partido	49	49,00	14	14,00	59	59,00
1	Tweets publicados por candidato	65	65,00	8	8,00	223	223,00
1	Tweets publicado por partido	99	99,00	0	0,00	22	22,00
3	Menciones a candidato (en @, en # y en texto)	506	1518,00	29	87,00	374	1122,00
5	Menciones a partido (en @, en # y en texto)	102	510,00	32	160,00	89	445,00
9	Menciones a hashtag de campaña	683	6147,00	2	18,00	180	1620,00
4	RT a candidatos	48	192,00	0	0,00	27	108,00
6	RT a partidos	34	204,00	0	0,00	6	36,00
2	Reply a candidatos	78	156,00	5	10,00	50	100,00
4	Reply a partidos	13	52,00	1	4,00	8	32,00
	SUMA	1729	9044	115	325	1094	3823

Fuente: elaboración propia

5.2.3.2. Resultados de la Investigación Elecciones Autonómicas a la Región de Murcia 2011

En las Elecciones Autonómicas a la Comunidad de Madrid 2011, como puede verse en la Tabla 42, tanto los resultados absolutos como los ponderados sólo encuentran valores aceptables para la candidatura del PSRM, siendo las desviaciones

²²⁴ Se ha descartado el hashtag “#votoutil” porque aunque fue muy utilizado por los simpatizantes de esta candidatura, también fue seguido por usuarios y plataformas asociados al movimiento #15M.

de IUV-RM e UPYD Murcia superiores al 6%. Dichas desviaciones, aunque, mejoran sensiblemente para UPYD Murcia tras aplicar la ponderación, continúan manteniendo valores escasamente estimables. Los resultados electorales se han tomado de la página Web de la Comunidad Autónoma de Murcia, http://resultadoselecciones2011.carm.es/15AU/DAU15999CM_L1.htm (Consultada el 22 del 06 de 2011). Una vez y más y como viene ocurriendo en las elecciones anteriormente analizadas, la candidatura correspondiente a IU aparece infradimensionada y la candidatura de UPYD sobredimensionada.

Tabla 42. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011

Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011							
	Suma Ponderada	%Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Twitter	Diferencia Sin Ponderación
PSRM	9044	24,82	-0,94	23,88	1729	21,31	2,57
IUV-RM	325	0,89	6,94	7,83	115	1,42	6,41
UPYD Murcia	3823	10,49	-5,99	4,5	1094	13,48	-8,98
SUMA	13192	36,21		36,21	2938	36,21	

Fuente: elaboración propia

Continuando con el método aplicado en los análisis de las Elecciones Generales 2011 y de las Elecciones Autonómicas a la Comunidad de Madrid se procede a repetir el cálculo teniendo en cuenta las variables que representan las menciones a candidatos y las menciones a partidos, con el fin de comprobar si los resultados mejoran tomando únicamente en cuenta las menciones, tal y como podría deducirse a tenor del estudio de investigación de Tumasjan et al., (2010) ya mencionado. Como puede observarse en la Tabla 44 y al igual que ocurría en las Elecciones Autonómicas a la Comunidad de Madrid, los resultados empeoran sustancialmente en dos de las opciones políticas (PSRM e UPYD) disminuyendo ínfimamente las desviaciones para la candidatura de IU-V-RM.

Tabla 43. Datos globales Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Sólo menciones

Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 - Sólo Menciones							
Ponderación	Variables	PSRM		IUUV-RM		UPYD Murcia	
		Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado
3	Menciones a candidato (en @, en # y en texto)	506	1518	29	87	374	1122
5	Menciones a partido (en @, en # y en texto)	102	510	32	160	89	445
	SUMA	608	2028	61	247	463	1567

Fuente: elaboración propia

Tabla 44. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Sólo menciones

Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 - Sólo Menciones							
	Suma Ponderada	%Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSRM	2028	19,11	4,77	23,88	608	19,45	4,43
IUUV-RM	247	2,33	5,50	7,83	61	1,95	5,88
UPYD Murcia	1567	14,77	-10,27	4,5	463	14,81	-10,31
SUMA	3842	36,21		36,21	1132	36,21	

Fuente: elaboración propia

Para continuar con el método aplicado en anteriores elecciones analizadas, replicamos el cálculo inicial, en el que se toman en cuenta todas las variables²²⁵, pero aplicándolas únicamente a los candidatos. Como puede verse en la Tabla 45, los resultados, con o sin ponderación, continúan empeorando salvo para la candidatura IUUV-RM que mantiene valores similares.

Tabla 45. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Candidato

Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 - Candidatos							
	Suma Ponderada	%Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSRM	1983	19,30	4,58	23,88	749	17,55	6,33
IUUV-RM	129	1,26	6,57	7,83	66	1,55	6,28
UPYD Murcia	1609	15,66	-11,16	4,5	730	17,11	-12,61
SUMA	3721	36,21		36,21	1545	36,21	

Fuente: elaboración propia

²²⁵ No se incluyen menciones a *hashtags* de campaña.

Aplicando la opción de “sólo menciones” para el perfil “candidato”²²⁶, los resultados mejoran en comparación con la toma de todas las actividades para candidato (Tabla 46) y muestran números más aproximados a los obtenidos con el cálculo inicial.

Tabla 46. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Candidatos – Sólo menciones

Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 - Candidatos - Sólo Menciones				
	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSRM	23,88	506	20,16	3,72
IUV-RM	7,83	29	1,16	6,67
UPYD Murcia	4,5	374	14,90	-10,40
SUMA	36,21	909	36,21	

Fuente: elaboración propia

A continuación, repetimos los cálculos referenciados para la opción única de “partidos”. Tal y como puede observarse en las tablas referidas a “candidatos”, los resultados sin ponderación que se muestran en las Tablas 47 y 48 se asemejan más a los obtenidos en el cálculo inicial que toma en cuenta tanto a candidatos como a partidos, que los obtenidos para la opción única “candidatos”. En los valores tras aplicar la ponderación, no se observan mejorías sino que empeoran comparando con aquellos en su misma situación.

Tabla 47. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Partidos

Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 - Partidos							
	Suma Ponderada	%Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSRM	914	19,63	4,25	23,88	297	20,37	3,51
IUV-RM	178	3,82	4,01	7,83	47	3,22	4,61
UPYD Murcia	594	12,76	-8,26	4,5	184	12,62	-8,12
SUMA	1686	36,21		36,21	528	36,21	

Fuente: elaboración propia

²²⁶ En este caso, al ser una única la variable analizada, no se calculan los valores con ponderación al resultar inútiles (matemáticamente, arrojan idéntico resultado que “sin ponderación”)

Este resultado, en disparidad con lo encontrado en elecciones anteriores, se hace más a los resultados obtenidos por Tumasjam et al. (2010) acerca de las elecciones alemanas.

Nuevamente en estas elecciones, al igual que en las Elecciones Generales 2011, tal y como puede comprobarse en Tabla 48, tampoco la toma en cuenta única de “menciones a partido” mejora absolutamente todos los resultados, encontrándose desviaciones similares o mayores que en los casos propuestos anteriormente (general y candidato). En esta ocasión, se mantiene la opción de UPYD Madrid, mientras que se incrementan los desajustes para el resto de opciones políticas.

Tabla 48. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Partidos – Sólo menciones

Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 - Partidos - Sólo Menciones				
	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSRM	23,88	102	16,56	7,32
IUV-RM	7,83	32	5,20	2,63
UPYD Murcia	4,5	89	14,45	-9,95
SUMA	36,21	223	36,21	

Fuente: elaboración propia

5.2.3.3. Conclusiones Análisis Twitter en las Elecciones Autonómicas a la Región de Murcia 2011

También en el caso de las Elecciones Autonómicas a la Región de Murcia 2011, monitorizar sólo partidos o sólo candidatos ofrece diferentes resultados que mejoran o empeoran en cada caso dependiendo de la opción política monitorizada, si bien la tendencia general es al empeoramiento. Así, las únicas mejoras encontradas se han encontrado de forma ligera (+0,86%) en el partido UPYD y con un mayor calado (+1,80%) para la opción IUV-RM (Izquierda Unida en esta región). Cuando se han contabilizado únicamente las actividades para Candidato, los resultados empeoran notablemente en todos los casos salvo para, una vez más, IUV-RM que prácticamente se mantiene (+0,13%).

Podemos concluir que en las Elecciones Autonómicas a la Región de Murcia 2011, los resultados mejoran o empeoran en mayor o menor medida para una u otra opción política dependiendo de si se mide para Candidato o para Partido, siendo en general el mejor resultado cuando se toman en cuenta ambas posibilidades.

Cuando se toman en cuenta sólo las menciones, ocurre de forma similar, en general los resultados empeoran, mejorando para una y otra opción según se consideren para Partido o para Candidato.

5.2.3.4. Encuestas Electorales versus Análisis Twitter en las Elecciones Autonómicas a la Región de Murcia 2011

Con el fin de establecer una comparación entre la posible capacidad de eficacia de este método en la predicción de resultados electorales y la eficacia demostrada, en mayor o menor medida, por las encuestas o métodos tradicionales, se han recogido los resultados de las encuestas preelectorales publicadas por los medios de comunicación y la encuesta preelectoral sondeada por el CIS.

Los resultados de las encuestas se han obtenido en las siguientes direcciones Web²²⁷:

CIS ²²⁸	http://datos.cis.es/pdf/Es2880mar_A.pdf
SYNOVATE	http://www.antena3.com/especiales/noticias/elecciones-autonomicas-municipales/lograria-mayoria-absoluta-comunidad-madrid-sevilla_2011050202145.html
CEMOP	http://www.cemopmurcia.es/encuestas/informe-preelectoral-region-de-murcia-elecciones-autonomicas-2011/

²²⁷ Consultadas el 05 de abril de 2012.

²²⁸ PREELECTORAL ELECCIONES.AUTONÓMICAS Y MUNICIPALES, 2011 (Región de Murcia)

Tanto los resultados de la encuestas como los obtenidos en el análisis de Twitter, se expresan en “% de intención de voto”. Como muestra de resultados Twitter, se han tomado los obtenidos después de aplicar la ponderación.

En la primera tabla (Tabla 49) se indican los valores obtenidos por cada una de las encuestas para cada una de las opciones políticas estudiadas. En la siguiente tabla (Tabla 50), se muestran las diferencias absolutas para cada una de ellas. Una vez, y recalando que en este análisis falta la candidatura ganadora, las diferencias entre partidos se repiten en todos los espacios analizados:

- La opción PSRM (PSOE de la Región de Murcia) se mantiene fiel a los resultados e incluso mejora en mucho las expectativas de las encuestas.
- La candidatura de IUVRM (coalición Izquierda Unida-Los verdes de la Región de Murcia) aparece infravalorada.
- UPYD Murcia (UPYD de la Región de Murcia) aparece sobrevalorada.

Cabe decir frente a la obviedad que salvo en uno de los casos, los resultados que arroja el análisis de Twitter, por sí sólo, no puede sustituir a la predicción obtenida por las encuestas.

El MAE conseguido en Twitter es de un 4,63% y por tanto, sin valor predictivo al superar los rangos aceptados en las encuestas de opinión pública.

Tabla 49. Pronósticos encuestas versus resultados Twitter en las Elecciones Autonómicas a la Región de Murcia 2011²²⁹

Elecciones Región de Murcia 2011 - Encuestas versus Resultado Ponderado Twitter				
Pronósticos Encuestas				
	CIS	CEMOP	Twitter	Resultados
PSRM	26	25,9	24,82	23,88
IUV-RM	6,6	8,5	0,89	7,83
UPYD Murcia	2,3	3,6	10,49	4,5

Fuente: elaboración propia

Tabla 50. MAE pronósticos encuestas versus resultados Twitter en las Elecciones Autonómicas a la Región de Murcia 2011

Elecciones Región de Murcia 2011 - Encuestas versus Resultado Ponderado Twitter				
Diferencia Asoluta				
	CIS	CEMOP	Twitter	Resultados
PSRM	2,12	2,02	0,94	23,88
IUV-RM	1,23	0,67	6,94	7,83
UPYD Murcia	2,20	0,90	5,99	4,5
% MAE	1,85	1,20	4,63	

Fuente: elaboración propia

²²⁹ En Tablas 49 y 50, todas las cifras expresan porcentajes.

5.2.4. Elecciones al Parlamento de Cataluña 2010

5.2.4.1. *Datos de Partida Elecciones al Parlamento de Cataluña 2010*

Las Elecciones al Parlamento de Cataluña 2010 fueron las primeras que se monitorizaron, cuando en ese año 2010 las investigadoras Mariluz Congosto de la Universidad Carlos III de Madrid y la autora de este trabajo se cruzaron en sus distintas investigaciones e iniciaron una andadura conjunta. Desde entonces, se ha perfeccionado la metodología de seguimiento. Es necesario manifestar que, dada la velocidad a la que Internet se desarrolla, el número de usuarios de la red Twitter en España se ha duplicado en sólo un año (intervalo 2010-2011), por lo que el análisis de este proceso electoral debe tomarse como iniciador y sus conclusiones más como referencia que como acto final.

Sumado a las objeciones anteriores, cabe sumar que de esta ocasión, sólo se tiene consciencia de la participación referida a partidos, no habiéndose seguido, por las razones de iniciación metodológica mencionadas, a los candidatos en campaña²³⁰. No obstante el seguimiento de esta campaña electoral resulta significativo en un contexto autonómico por el número de opciones políticas que se vigilan, el cuál supera en mucho al de resto de situaciones regionales estudiadas en esta investigación, lo cual puede ayudar a comprender como suceden los acontecimientos en este entorno y/o indicar diferencias de peso al compararse con otros procesos electorales con cualidades y cantidades políticas menores.

Los resultados electorales se han tomado de la página Web de la Generalitat de Catalunya, http://www.gencat.cat/governacio/eleccions/eleccions2010/resultats2010/09AU/DAU09999CM_L2.htm. (Consultada el 14 del 12 de 2010).

²³⁰ (alguno de los cuales nos consta que no mantuvo perfil alguno en Twitter, como es el caso de Artur Mas, candidato de la opción política CIU-Convergencia y Unió).

La Tabla 52 recoge la suma del número diario de acciones Twitter (variables) por opción política desde el día 12 del 11 de 2010 hasta el día 27 del mismo mes. Las opciones políticas que se han monitorizado, como ya se ha indicado referentes únicamente a partidos políticos y hashtags de campaña, como se refleja en la Tabla 51, son: CiU (Convergencia i Unió), PSC-PSOE (Partido Socialista de Cataluña), PP Catalunya (Partido Popular de Cataluña), ICV-EUiA (Iniciativa per Catalunya Verds - Esquerra Unida i Alternativa), ERC-ESQUERRA (Esquerra Republicana de Catalunya), SI (Solidaritat Catalana per la Independència), y C's (Ciudadanos - Partido de la Ciudadanía). Los valores que se recogen por cada una de ellas son “absoluto” (suma de los valores obtenidos diariamente) y “ponderado” (resultado de aplicar la ponderación indicada al valor absoluto obtenido).

El volumen de actividad o número de *tweets* es de 72.485.

Los términos monitorizados para cada partido fueron los siguientes:

Tabla 51. Relación de términos monitorizados en las Elecciones al Parlamento de Cataluña 2010

	Partido	Menciones a partido	Hashtag de campaña
CiU	@ciu	@ciu	#unacatalunyamillor
PSC	@socialistes_cat	@socialistes_cat @psc	#garantiadeprogres
PP	@PPCatalunya	@PPCatalunya	#solucions
ICV	@icveuia	@icveuia	#joanherrera
ERC	@Esquerra_ERC	@Esquerra_ERC Esquerra erc #esquerra #erc	#gentvalenta ²³¹

²³¹ Se omiten los hashtag #votaindependencia por haberse constatado que es de uso común para las candidaturas de ERC y SI.

SI	@solidaritatcat	@solidaritatcat	#solidaritatcatalana
C's	@Ciutadans_Cs	@Ciutadans_Cs	#rebelate

Fuente: elaboración propia

Tabla 52. Datos globales Twitter en las Elecciones al Parlamento de Cataluña 2010

Twitter en las Elecciones al Parlamento de Cataluña 2010													
Ponderación	Variables	CIU		PSC		PP Cataluña		ICV-EUIA		ERC-ESQUERRA		SI	
		Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado	Absoluto	Ponderado
1	Followers partido	295	295	199	199	117	117	194	194	179	179	248	248
1	Tweets publicado por partido	529	529	607	607	232	232	381	381	193	193	377	377
5	Menciones a partido (en @, en # y texto)	5076	25380	2523	12615	649	3245	1171	5855	803	4015	410	2050
9	Menciones a hashtag de campaña	5466	49194	3559	32031	151	1359	1002	9018	1325	11925	936	8424
6	RT a partidos	362	2172	402	2412	70	420	139	834	145	870	134	804
4	Reply a partidos	203	812	167	668	56	224	45	180	41	164	12	48
	SUMA	11931	78382	7457	48532	1275	5597	2932	16462	2686	17346	2117	11951

Fuente: elaboración propia

5.2.4.2. Resultados de la Investigación Elecciones al Parlamento de Cataluña 2010

Como puede observarse en la Tabla 53, las desviaciones con respecto los resultados electorales definitivos son sostenibles salvo para la candidatura del Partido Popular de Cataluña (PP Cataluña), en cuyo caso la diferencia es más que significativa. Para el resto, los valores resultantes del análisis Twitter pueden considerarse acertados ya que se encuentran en el rango 2%-3% que se espera de las encuestas. Cuando se aplica la ponderación, los valores mejoran en todas las candidaturas, excepto nuevamente en el Partido Popular, para la que la diferencia se incrementa en un punto porcentual.

Tabla 53. Resultado Twitter en las Elecciones al Parlamento de Cataluña 2010

Resultado Twitter en las Elecciones al Parlamento de Cataluña 2010							
	Suma Ponderada	%Suma Ponderada	Diferencia Con Ponderación	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
CIU	78382	37,69	0,74	38,43	11931	35,59	2,84
PSC	48532	23,34	-4,96	18,38	7457	22,24	-3,86
PP Cataluña	5597	2,69	9,68	12,37	1275	3,80	8,57
ICV-EUIA	16462	7,92	-0,55	7,37	2932	8,75	-1,38
ERC-ESQUERRA	17346	8,34	-1,34	7	2686	8,01	-1,01
SI	11951	5,75	-2,46	3,29	2117	6,31	-3,02
C's	9353	4,50	-1,11	3,39	1854	5,53	-2,14
SUMA	187623	90,23		90,23	30252	90,23	

Fuente: elaboración propia

Continuando con los cálculos únicos aplicados a menciones aplicados en las elecciones anteriormente analizadas, repetimos el método teniendo en cuenta únicamente la variable “menciones”. La Tabla 54 muestra los resultados que se obtienen. Tomando en conjunto el total de las candidaturas, los valores mejoran al tomar en cuenta únicamente esta variable, ya que sólo se incrementa la diferencia del valor que tenemos con Twitter comparado con los resultados electorales para la candidatura ICV-EUiA, mejorando en mayor o menos medida en el resto de ellas.

Tabla 54. Resultado Twitter en las Elecciones al Parlamento de Cataluña 2010 – Sólo menciones

Resultado Twitter en las Elecciones al Parlamento de Cataluña 2010 - Sólo Menciones				
	Resultado Electoral	Suma Twitter	% Twitter	Diferencia sin ponderación
CiU	38,43	5076	40,64	-2,21
PSC-PSOE	18,38	2523	20,20	-1,82
PP Cataluña	12,37	649	5,20	7,17
ICV-EUiA	7,37	1171	9,38	-2,01
ERC-ESQUERRA	7	803	6,43	0,57
SI	3,29	410	3,28	0,01
C's	3,39	637	5,10	-1,71
SUMA	90,23	11269	90,23	

Fuente: elaboración propia

5.2.4.3. Conclusiones Análisis Twitter en las Elecciones al Parlamento de Cataluña 2010

En el caso de las Elecciones al Parlamento de Cataluña 2010, en la que dadas las características excepcionales de su captura ya explicadas, se observa un buen resultado de la actividad en Twitter, similar en calidad al obtenido en las Elecciones Generales 2011. Los valores conseguidos para la opción Partido son bastante aceptables, rondando valores de diferencia que apenas superan el 3% admitido para cualquier encuesta de opinión tradicional, excepto para el Partido Popular de esta región, cuya diferencia es de un poco o nada aceptable 8,57%.

También en esta elección dado que se refiere a la actividad destinada a Partido, los resultados mejoran cuando se toman en cuenta únicamente las

menciones, puesto que todos los valores mejoran salvo un ligero empeoramiento de un 0,65% observado en la opción ICV-EUIA.

Es preciso recordar en esta elección que la introducción en el análisis de un gran número de partidos haya sido la causa que ha provocado los buenos resultados obtenidos aún disponiendo únicamente de los datos referidos a partidos. Dicha hipótesis coincide con las variaciones encontradas por Jungherr et al. (2011) en su trabajo sobre las elecciones alemanas de 2009, cuyos resultados difieren de los concluidos por Tumasjan et al. (2010) para el mismo proceso electoral, y en el que recordamos se manifiesta que la sola inclusión de los datos del Partido Pirata en los cálculos, arroja conclusiones distintas.

5.2.4.4. Encuestas Electorales versus Análisis Twitter en las Elecciones al Parlamento de Cataluña 2010

Una vez más, con el objeto de establecer una comparación entre la posible capacidad de eficacia de este método en la predicción de resultados electorales y la eficacia demostrada, en mayor o menor medida, por las encuestas o métodos tradicionales, se han recogido los resultados de las encuestas preelectorales publicadas por los medios de comunicación y la encuesta preelectoral sondeada por el CIS²³².

Los resultados de las encuestas se han obtenido en las siguientes direcciones Web²³³:

CIS ²³⁴	http://www.cis.es/cis/opencms/-
--------------------	---

²³² No se han tenido en cuenta las encuestas publicadas por la cadena TV3 (realizada por Ipsos - <http://www.324.cat/noticia/961126/politica/El-sondeig-de-TV3-i-Catalunya-Radio-pronostica-una-victoria-de-CiU-a-prop-de-la-majoria-absoluta>) y la cadena de radio Cope (GAD - <http://www.electometro.es/2010/11/ciu-ganaria-y-si-podria-tener-representacion-en-el-parlament-cope/>) ya que los datos que ofrecían y se han encontrado venían expresados en escaños y no en porcentaje de intención de voto.

²³³ Consultadas el 05/04/2012.

²³⁴ CIS. Preelectoral de Cataluña, Elecciones Autonómicas, 2010.

	Archivos/Marginales/2840_2859/2852/Es2852Cat.pdf
GESOP	http://www.elperiodico.com/es/noticias/elecciones-28n/20101122/ciu-acerca-aun-mas-mayoria-absoluta/593793.shtml
NC Report	http://www.electometro.es/2010/11/ciu-se-ecuentra-a-6-escanos-de-la-mayoria-absoluta-la-razon/

Tanto los resultados de la encuestas como los obtenidos en el análisis de Twitter, se expresan en “% de intención de voto”. Como muestra de resultados Twitter, se han tomado los obtenidos después de aplicar la ponderación.

En la primera tabla (Tabla 55) se indican los valores obtenidos por cada una de las encuestas para cada una de las opciones políticas estudiadas. En la siguiente tabla (Tabla 56), se muestran las diferencias absolutas para cada una de ellas. Como cabía esperar, a razón de los datos que muestran las Tablas 33 y 34, los porcentajes de intención de voto que se han obtenido monitorizando la actividad de campaña en la red Twitter son comparables a los obtenidos en las encuestas de opinión desempeñadas por diversos medios, salvo para el Partido Popular de Cataluña, en el que la diferencia es más que significativa. Se hace constar que la actividad de este partido en la red fue prácticamente nula o inexistente, realidad que podría enmascarar la actividad de los posibles votantes quienes se hubieran desvelado al encontrar un candidato al que apoyar en este canal.

El MAE de Twitter es de un 2,98%, que si bien es peor que el mostrado por todas las encuestas, se mantiene dentro de los márgenes aceptables de predicción de la opinión pública.

Tabla 55. Pronósticos encuestas versus resultados Twitter en las Elecciones al Parlamento de Cataluña 2010²³⁵

Elecciones al Parlamento de Cataluña 2010 - Encuestas versus Resultado Ponderado Twitter					
Pronósticos Encuestas					
CANDIDATURAS	CIS	GESOP	NC Report	Twitter	Resultados
CIU	38	39,1	39,4	37,69	38,43
PSC-PSOE	22,7	20	21,8	23,34	18,38
PP	9,7	10,3	11,4	2,69	12,37
ICV-EUiA	8,2	8	8,5	7,92	7,37
ERC-Esquerra	10,2	7,5	9,3	8,34	7
C's	3,5	4,3	3,2	4,50	3,39
SI	1	3,2	1,8	5,75	3,29
Periódicos		Periódico de Cataluña	La Razón		

Fuente: elaboración propia

Tabla 56. MAE pronósticos encuestas versus resultados Twitter en las Elecciones al Parlamento de Cataluña 2010

Elecciones al Parlamento de Cataluña 2010 - Encuestas versus Resultado Ponderado Twitter					
Diferencia Absoluta					
CANDIDATURAS	CIS	GESOP	NC Report	Twitter	Resultados
CIU	0,43	0,67	0,97	0,74	38,43
PSC-PSOE	4,32	1,62	3,42	4,96	18,38
PP	2,67	2,07	0,97	9,68	12,37
ICV-EUiA	0,83	0,63	1,13	0,55	7,37
ERC-Esquerra	3,2	0,5	2,3	1,34	7
C's	0,11	0,91	0,19	1,11	3,39
SI	2,29	0,09	1,49	2,46	3,29
% MAE	1,98	0,93	1,50	2,98	

Fuente: elaboración propia

²³⁵ En Tablas 55 y 56, todas las cifras expresan porcentajes.

5.3. Una Aproximación Metodológica: Resultado del Análisis según Tercera Línea de Investigación

Se hace patente que en esta tercera línea de investigación intentamos una aproximación a diversos estudios anteriores y de diversas disciplinas (Desart y Holbrook (2003) y (Jungherr et al. (2011), en los que se asegura que las fechas en las que los datos son recolectados, influyen positiva o negativamente en los resultados, adquiriendo, según los primeros, mayor certeza cuanto más próximos se encuentren de la jornada electoral, e incrementa la desviación según sugieren los segundos.

También se hace recordar que en esta ocasión se han tomado como referencia solamente dos procesos electorales, Elecciones Generales 2011 y Elecciones Autonómicas a la Comunidad de Madrid 2011, por las razones argumentadas previamente en este trabajo. Así mismo, sólo se muestran valores sin ponderación, puesto que los datos obtenidos con la misma son objeto de la línea de investigación anterior, además de ser débiles las mejoras obtenidas y que por tanto, no ampliarían en esta ocasión las conclusiones de lo que se muestra.

Se incluyen a continuación las tablas con los resultados para ambas elecciones. En las primeras se muestran los valores absolutos obtenidos a nivel de actividad Twitter y las segundas, muestran, como a lo largo de esta investigación, lo que esta actividad significa en porcentaje de voto en comparación con los resultados electorales definitivos.

Tabla 57. Datos globales Twitter en las Elecciones Generales 2011 – 7 días

Twitter en las Elecciones Generales 2011 - 7 días				
Variables	PSOE	PP	IU	UPYD
Followers candidato	8340	14104	1786	0
Followers partido	1131	5423	813	2254
Tweets publicados por candidato	738	431	54	0
Tweets publicado por partido	106	117	677	883
Menciones a candidato (en @, en # y en texto)	133854	233635	15645	19346
Menciones a partido (en @, en # y en texto)	33281	67246	18143	21364
Menciones a hashtag de campaña	35	32	796	3388
RT a candidatos	7563	5009	988	0
RT a partidos	501	1480	3357	3823
Reply a candidatos	9455	10004	700	0
Reply a partidos	556	1011	796	1928
	195560	338492	43755	52986

Fuente: elaboración propia

Tabla 58. Resultados Twitter en las Elecciones Generales 2011 – 7 días

Resultado Twitter en las Elecciones Generales 2011 - 7 días				
	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSOE	28,73	195560	26,34	-2,39
PP	44,62	338492	45,59	0,97
IU-LV	6,92	43755	5,89	-1,03
UPYD	4,69	52986	7,14	2,45
SUMA	84,96	630793	84,96	

Fuente: elaboración propia

Cómo puede observarse, los resultados para Elecciones Generales 2011 son mejores para todas las candidaturas excepto para la opción política UPYD, en la que su incremento de obtención votos parece revelarse como efecto de que en los últimos días de campaña, la actividad de sus seguidores se intensificó, quizás debido al efecto que ya se ha mencionado sobre que los partidos minoritarios refuerzan su posicionamiento en este tipo de canales sabiéndose “olvidados” por los medios de comunicación tradicionales.

Insertamos a continuación, las tablas sobre las Elecciones Autonómicas a la Comunidad de Madrid 2011, que se incorporan con el ánimo de refutar o no la mejoría encontrada en las Elecciones Generales 2011 al tomar los datos de siete días antes de las elecciones propiamente dichas.

Tabla 59. Datos globales Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011 – 7 días

Twitter en las Elecciones Autonómicas a la CAM 2011 - 7 días				
Variables	PSM	PP Madrid	IU-LV	UPYD Madrid
Followers candidato	145	801	150	66
Followers partido	52	385	43	70
Tweets publicados por candidato	727	54	6	64
Tweets publicado por partido	81	407	56	299
Menciones a candidato (en @, en # y en texto)	9318	15833	338	977
Menciones a partido (en @, en # y en texto)	519	4402	115	292
Menciones a hashtag de campaña	973	2216	139	211
RT a candidatos	549	289	3	10
RT a partidos	38	722	52	88
Reply a candidatos	542	1680	20	87
Reply a partidos	53	523	10	36
SUMA	12997	27312	932	2200

Fuente: elaboración propia

Tabla 60. Resultados Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011 – 7 días

Resultado Twitter en las Elecciones Autonómicas a la CAM 2011 - 7 días				
	Resultado Electoral	Suma Twitter	% Suma Twitter	Diferencia Sin Ponderación
PSM	26,27	12997,24	28,10	-1,83
PP Madrid	51,73	27312	59,05	-7,32
IU-LV	9,61	932	2,02	7,59
UPYD Madrid	6,32	2200	4,76	1,56
SUMA	93,93	43441,24	93,93	

Fuente: elaboración propia

En las Elecciones Autonómicas a la Comunidad de Madrid 2011, como puede verse en las Tablas 59 y 60, parecería que los resultados que mejoran para uno

empeoran para otro. El resultado, en conclusión, es que la toma de datos reducida a los siete últimos días de campaña no mejora los resultados iniciales desvelados en las anteriores líneas de investigación.

Es por tanto factible afirmar que no sólo la toma de datos, sino el resto de condicionantes y variables que acontecen en un evento electoral, son responsables de los resultados que puedan obtenerse a partir de la actividad Twitter monitorizada en ellos. Además de las fechas de captura, factores intrínsecos como el tamaño de la población, el volumen de tweets, la participación de los candidatos, y el número de sujetos monitorizados, junto con factores extrínsecos al propio análisis pero dados en una elección, como la situación o no de incumbencia de sus candidatos o las características sociodemográficas del entorno, podrían afectar en mayor o menor medida a los resultados de la actividad Twitter.

5.4. Twitter y el resultado electoral: posibles causas de eficacia y desviación

Volvemos a retomar el peso que en este tipo de investigaciones puede llegar a alcanzar el tamaño de la muestra. Debe entenderse por muestra, no sólo el porcentaje de población con participación en Twitter sino, y de forma esencial, el volumen de la actividad de ésta. En este sentido, la actividad constatada en cada uno de los procesos electorales, entendida ésta como volumen o número de *tweets*, ha sido la siguiente:

Tabla 61. Volumen de tweets en los procesos electorales monitorizados

Elección	Número de tweets
Elecciones Generales 2011	1.592.872
Elecciones Autonómicas a la Comunidad de Madrid 2011	134.193
Elecciones Autonómicas a la Región de Murcia 2011	3.303

Elecciones al Parlamento de Cataluña 2010 ²³⁶	72.485
--	--------

Es constatable que las diferencias en cantidad entre una elección general y una elección regional, son considerables, lo cual nos anima a considerar que pueda ser necesaria una cantidad de datos o *tweets* proporcional al universo poblacional de dicha elección, tal y como se concluyó en el estudio de Teewtminster (2010) en el que se afirmaba que la correlación es mayor cuando el volumen de *tweets* de una elección es considerable, y a más, las características demográficas no son determinantes en estos casos.

Respecto al porcentaje de población usuaria de esta Red, las conclusiones que pueden obtenerse son similares: a mayor población representada, mayor marco muestral posible y mejores resultados obtenidos. Una vez más, las Elecciones al Parlamento de Cataluña 2010 destacan por no ajustarse a estas reglas, dado que tanto su marco muestral como su volumen de datos recogidos (número de *tweets*) es menor que en el caso de Elecciones Autonómicas a la Comunidad de Madrid 2011. Se retoma la posibilidad de que el mayor número de partidos representados, haya sido en esta ocasión el factor relevante del buen resultado obtenido. Replicamos a continuación los números de las poblaciones usuarias de Twitter en cada uno de los procesos electorales.

Tabla 62. Porcentaje de población con derecho a voto usuaria de la red Twitter

Elección	Porcentaje de población
Elecciones Generales 2011	30,13%
Elecciones Autonómicas a la Comunidad de Madrid 2011	27,10%
Elecciones Autonómicas a la Región de Murcia 2011	0,82%

²³⁶ Debe tomarse en cuenta que sólo se monitorizaron partidos, por lo que el volumen de datos pudo haber sido mayor.

Elecciones al Parlamento de Cataluña 2010	19,69%
---	--------

Parece así manifestarse una relación entre la capacidad de Twitter en el análisis electoral y, tanto el número de tweets capturados en una elección dada, como el número de opciones electorales (candidatos y partidos) introducidas en el análisis, no guardando ninguna de ellas (volumen de datos y unidades monitorizadas) relación directa con los tamaños de población y marco muestral. Para intentar comprenderlo se ha construido la siguiente tabla de la que puede desprenderse que el porcentaje de usuarios de Twitter no influye directamente en el ajuste de la predicción, siendo el número de tweets determinante en la veracidad de los resultados obtenidos, conclusión que coincide con la informada por Tweetminster (2010): “A mayor volumen de datos, mayor grado de acierto. El modelo se comporta mejor en los ámbitos nacional (2.010.000 tweets) y regional (con una media de 37,000 tweets) que a nivel local (con una media de 667 tweets por circunscripción).”

Por otro lado, teniendo en cuenta según la tabla que se inserta, el comportamiento de las Elecciones al Parlamento de Cataluña 2010 es excepcional en cuanto que aún disponiendo de menor número de tweets que las Elecciones Autonómicas a la Comunidad de Madrid 2011 y siendo también menor el porcentaje de la población con derecho a voto que es usuaria de Twitter (19,69% frente a 27,70), consiguió un MAE menor. La explicación parece encontrarse en el hecho de que el número de partidos políticos tomados en consideración es de prácticamente el doble, y que se apoya en los estudios de Jungherr et al. (2011) y en los resultados sobre Elecciones Generales 2011 presentados por Congosto y Aragón (2012) en el I Congreso de ALICE²³⁷, en los que incluyendo una opción política más que en la presente investigación (Equo), se obtuvieron mejores valores²³⁸.

²³⁷ <http://www.alice-comunicacionpolitica.com/index.php>

²³⁸ No obstante es preciso que se tenga en cuenta que aún disponiendo de los mismos dataset, se difiere en el periodo de captura calculado y en las variables que se manejan.

Continuando con el ánimo de encontrar una relación de tipo causal entre el MAE que arroja los resultados de Twitter y los datos referidos a participación y población, en la siguiente tabla se incluyen las cifras obtenidas relacionando esta vez a los usuarios únicos²³⁹ que participaron en esta red. Como puede extraerse, las relaciones que se establecen entre “usuarios únicos” y “usuarios twitter”, y entre aquellos y el “volumen o número de tweets”, mantienen una relación directa con la fiabilidad del carácter predictivo obtenido en Twitter. Es por tanto, que a tenor de estas cifras, nos atrevemos a sugerir que a medida que el valor de ambas relaciones se aleja de cero, la validez de los resultados que nos ofrece el análisis de Twitter se encuentra más alejado de la realidad, y viceversa. De hecho, aunque tanto el número de tweets como el número de usuarios de la red es mayor en las Elecciones Autonómicas a la Comunidad de Madrid 2011 que en las Elecciones al Parlamento de Cataluña 2010, los resultados de su análisis presentan un resultado bastante mejor (MAE 4,10% frente a MAE 2,98%). Este comportamiento puede explicarse con el siguiente razonamiento: una cifra menor de “Usuarios únicos/Volumen de tweets” indicaría que los tweets han sido publicados por un mayor número de individuos únicos y por tanto, la muestra es mayor y diversa; a medida que la relación incrementa su valor, nos encontraríamos con que muchos de los tweets recogidos podrían haber sido publicados por un número determinado de usuarios muy activos políticamente con lo que esos tweets resultan menos objetivos para el resultado de la investigación. De hecho, estas afirmaciones se ven refrendadas por las conclusiones dictadas por Barberá y Rivero (2012) en su análisis de las Elecciones Generales 2011: *“Es más, nuestros resultados también muestran la profunda desigualdad existente en el comportamiento de los grupos que se encuentran sobre-representados en Twitter: los seguidores de los partidos políticos son mucho más activos en esta conversación que el resto de usuarios.”* (Barberá y Rivero, 2012, p.19)

²³⁹ El total se obtiene de la suma de los usuarios únicos que se dieron cada uno de los quince días monitorizados.

Tabla 63. Relación del MAE obtenido en Twitter con el volumen de tweets por población y usuarios

Relación del MAE obtenido en Twitter con el volumen de tweets por población y usuarios						
	Población con derecho a voto	Usuarios Twitter	Volumen de tweets	MAE	Tweets/Población derecho a voto	Tweets/Usuarios Twitter
Generales 2011	26.177.427	7.888.000	1.592.872	1,57%	0,06	0,20
Parlamento de Cataluña 2010	3.964.042	780.423	72.485	2,98%	0,02	0,09
Comunidad de Madrid 2011	3.538.353	980.267	134.193	4,10%	0,04	0,14
Región de Murcia 2011	786.425	6.443	3.303	4,63%	0,004	0,51

Fuente: elaboración propia

Tabla 64. Relación del MAE obtenido en Twitter con el número de usuarios únicos por población y volumen de tweets

Relación del MAE obtenido en Twitter con el número de usuarios únicos por población y volumen de tweets								
	Población con derecho a voto	Usuarios Twitter	Volumen de tweets	Usuarios únicos	MAE	Usuarios únicos/Población derecho a voto	Usuarios únicos/Usuarios Twitter	Usuarios únicos/Volumen de tweets
Generales 2011	26.177.427	7.888.000	1.592.872	197.817	1,57%	0,008	0,025	0,12
Parlamento de Cataluña 2010	3.964.042	780.423	72.485	26.560	2,98%	0,007	0,034	0,37
Comunidad de Madrid 2011	3.538.353	980.267	134.193	63.117	4,10%	0,018	0,064	0,47
Región de Murcia 2011	786.425	6.443	3.303	1.702	4,63%	0,002	0,264	0,52

Fuente: elaboración propia

5.5. Conclusiones

En términos generales, sin duda los análisis para Elecciones Generales 2011 son los que han obtenido mejores resultados, encontrándose los mismos cercanos al resultado electoral definitivo y siendo las diferencias encontradas similares y, en varios casos, menores que las obtenidas por las encuestas de opinión. Dicha característica se observa también en las Elecciones al Parlamento de Cataluña 2010. En el resto de procesos, Elecciones Autonómicas a la Comunidad de Madrid 2011 y Elecciones Autonómicas a la Región de Murcia 2011, los resultados de la actividad Twitter presentan grandes diferencias para varias de sus opciones políticas.

Las candidaturas que presentan mayores desviaciones en todos los casos son las correspondientes a IU (Izquierda Unida) y UPYD (Unión Progreso y Democracia). En el primer caso, la tendencia es a la infravaloración y en el segundo a la sobrevaloración. Entre las posibles causas podría argumentarse que IU es un

partido político que mantiene un voto cautivo cuya mayoría no está representada en esta Red al no participar masiva o proporcionalmente en ella. UPYD, por el contrario, es un partido joven, de reciente aparición, con un electorado joven y participe en Twitter. Está demostrado además que los partidos minoritarios tal como demostró en su estudio sobre las Ciber campañas en las elecciones de Australia 2004 de Gibson y McCallister (2005), muestran una mayor tendencia a hacer uso de Internet, dada la escasa capacidad que tienen de ser amplificadas por los medios de comunicación tradicionales como la prensa escrita y la televisión: *“Los partidos minoritarios se benefician de las páginas Web’s o ciber campañas al ser un espacio en el que poder emitir y ser escuchados. Espacio que les es “negado” por los medios de comunicación tradicionales.”* (Norris, 2003, p.43)

Es preciso señalar las particularidades intrínsecas a cada elección, así como aquellas que han afectado a las particularidades de su monitorización en Twitter, a saber:

- En las Elecciones Generales 2011, el tamaño de la muestra supera al resto de las elecciones estudiadas, tanto en porcentaje de población usuario de la Red, como en volumen de tweets y usuarios únicos.
- En las Elecciones al Parlamento de Cataluña 2010, el número de partidos u opciones políticas monitorizadas prácticamente dobla el número de cualquiera del resto de procesos analizados. Dados los buenos resultados obtenidos en esta elección, esta realidad podría en parte contradecir el trabajo de Jungherr et al. (2011) cuando afirman que la inclusión de nuevos partidos, como el Partido Pirata, puede modificar considerablemente las predicciones basadas en el análisis Twitter. Es preciso de nuevo recordar que se ha de tener en cuenta el peso que los partidos minoritarios alcanzan en Twitter, al ser una de sus plataformas Red accesibles para la difusión de sus propuestas, y que por tanto, su presencia en ella se encontraría sobredimensionada respecto su peso

real. Además, en esta ocasión sólo se encuentra monitorizada la actividad Twitter referida a Partido.

- En las Elecciones a la Región de Murcia 2010, la opción política ganadora, el Partido Popular, no mantuvo presencia alguna en esta red (ni el partido ni su candidato).

En este apartado, además de las conclusiones separadas mencionadas, se intentará obtener unas conclusiones generales a tenor de la observación comparada de lo que ha sucedido en los cuatro procesos electorales estudiados. Para ayudar a su mejor comprensión, dichas conclusiones se organizan entorno a los distintos análisis desempeñados:

- Resultado Twitter (Candidato+Partido)
 - Cuando se aplica la ponderación, las diferencias mejoran en todos los casos, y en aquellos en los que empeora, lo hace ligeramente con un $\pm 0,50\%$ ²⁴⁰ (PP Madrid e IUUV-RM).
 - La opción política que mantiene mejores y constantes resultados en todas las elecciones, es el PSOE (Partido Socialista Obrero Español).
 - Cuando el universo de la muestra disminuye, referido este al porcentaje de población que expresa pertenecer a la Red Twitter, las expectativas Twitter aunque empeoran en comparación con universos mayores, no lo hacen de forma progresiva, como puede deducirse del hecho de que los resultados de la actividad Twitter son mejores en las Elecciones al Parlamento de Cataluña 2010 que en las Elecciones Autonómicas a la Comunidad de Madrid 2011, siendo la población representada en Twitter mayor en esta última que en la anterior. Además, en las Elecciones

²⁴⁰ Tomando como referencia los resultados sin ponderación

Autonómicas a la Región de Murcia 2011, cuya población en Twitter es muchísimo menor que en las de la Comunidad de Madrid, los resultados no empeoran numéricamente en el mismo orden o porcentaje en que dicha población decrece (es preciso recordar aquí, que en la Región de Murcia tanto el partido como el candidato ganador y que se presentaba a su reelección, carecían de presencia en esta red).

- Resultado Twitter (Candidato+Partido) sólo Menciones.
 - Cuando se toman en cuenta sólo las menciones, sólo se observa mejoría para la opción PP Madrid.
- Resultado Twitter Candidato y Resultado Twitter Partido. Cuando se toman en cuenta las actividades sólo por Partido o sólo por Candidato, la tendencia general es a empeorar en todas las elecciones y para todas las opciones políticas, con las siguientes excepciones:
 - PP de Madrid mejora para Candidato y empeora notablemente para Partido.
 - IU (Izquierda Unida) mantiene valores similares en el caso de Partido en todas las elecciones analizadas
 - UPYD mantiene siempre valores similares para Partido.

En la línea de investigación cuyo objetivo trataba de dilucidar si el periodo establecido para la toma de datos afectaba a los resultados, se ha podido comprobar que son mejores en procesos de volumen de actividad considerable, como es el caso de las Elecciones Generales 2011 y se mantiene o empeora (dado que las desviaciones continúan en uno u otro sentido) en elecciones de menor envergadura.

Es destacable y merece tenerse en cuenta, que en el caso de las Elecciones al Parlamento de Cataluña 2010²⁴¹, los resultados son realmente favorables aún

²⁴¹ En su análisis, a diferencia del resto de elecciones, se ha incluido la variable “hashtag”. No obstante, se ha comprobado que la inclusión de ésta en los análisis referidos a la opción Partido, no mejora los resultados en el resto de procesos analizados.

teniendo sólo en consideración la actividad Twitter referida a Partido, actividad que en el resto de ocasiones se ha manifestado negativamente para la investigación. Bien pudiera ser que la confluencia de numerosos partidos en un mismo contexto (prácticamente el doble que en el resto de casos) haya favorecido el acierto.

Más allá de las variables y opciones políticas vigiladas, se han encontrado indicios reveladores acerca de las posibles causas de la eficacia o desviación en el análisis de Twitter como herramienta de predicción, referidas al volumen de datos manejados en relación con el volumen de la muestra. Los hallazgos han sido los siguientes:

- El volumen de tweets recolectados es de suma relevancia para lograr un buen resultado, coincidiendo con las conclusiones del trabajo elaborado por Twitminster (2010). Así, la precisión es muy buena en Elecciones Generales 2011, en la que se supera el millón de tweets tratados. Para ámbitos regionales, en los que dicho volumen disminuye considerablemente, no existe relación alguna entre el volumen de datos y las poblaciones de estudio.
- El número de partidos que se incluyen en el estudio puede afectar a los resultados. La introducción de partidos minoritarios o con escasa o nula representación parlamentaria, así como de partidos de reciente aparición, que tuvieron cierta presencia o viralidad en las redes sociales, puede modificar substancialmente el análisis en Twitter, tal y como advierten Jungherr et al. (2011) refiriéndose al Partido Pirata en las elecciones alemanas de 2009. En esta investigación, nos estaríamos refiriendo a Equo, SI y al propio UPYD.
- Se ha hallado una causalidad directa entre la eficacia del análisis electoral en Twitter y la relación entre “usuarios únicos” y “usuarios de Twitter” y “volumen de tweets”, de forma que cuanto más alejados de cero se encuentren los valores de estas relaciones menor

es la eficacia de su predicción, y viceversa. Dicha condición se cumple en las cuatro elecciones monitorizadas.

También se ha constatado que en aquellas opciones políticas en las que la actividad Twitter general ha sido muy reducida en comparación con su masa de posibles votantes, las diferencias comparativas con los resultados electorales son notables, siendo éstas siempre en carácter negativo u opción política infradimensionada (IU). De igual forma pero en sentido contrario, en aquellas opciones políticas en las que la actividad ha sido incrementalmente desproporcionada las diferencias, también notables, con los resultados electorales adquieren un carácter positivo u opción política sobredimensionada (UPYD).

Conclusiones

6. Conclusiones

Entre todas las conclusiones revisadas en diversos estudios, hemos fijado la atención en los aspectos que guardan una mayor relación con el objeto de nuestro trabajo, ya que nos ayudarían a entender cómo se transmite una idea o mensaje en una red social e intentar, posteriormente, de buscar similitudes o diferencias con las redes sociales que se configuran en Internet (Facebook y Twitter): el peso de la centralidad, determinación del grado de influencia, la fuerza de los vínculos débiles, la capacidad de contagio entre individuos y la amplitud de la cadena de difusión.

En la red social Facebook algunos autores han determinado que la influencia existe y que ésta funciona igual que en el mundo real. Sin embargo, otros estudios afirman que el modelo de difusión en Facebook difiere de los modelos clásicos en lo que se refiere al peso de los influyentes y a la longitud de las cadenas “infectadas”. Mientras que tradicionalmente se admitía que un evento comienza con un reducido número de nodos iniciadores que son capaces de afectar a los vecinos más vulnerables, en Facebook los contagios que se convierten en cascadas globales se dan en eventos que comienzan en un gran número de nodos que ponen en marcha un gran número de cadenas cortas, cada una de las cuales colisiona con rapidez en una sola estructura de gran tamaño. Por tanto, no se puede identificar a los iniciadores como lo más importante en un proceso de difusión cultural, ya que en redes sociales como Facebook una gran cantidad de individuos tienen la posibilidad de desencadenar una cascada global de gran tamaño. Además, aunque ahora todos podemos influir en los demás, ha surgido la aparición de una nueva clase de influyentes a los que denominamos “super-influyentes”. Individuos que gracias a los social media y a las herramientas que proveen para crear y compartir juicios, superan el grado medio de la influencia difundiendo su opinión más lejos y más ampliamente.

La red Twitter ha demostrado es una capacidad asombrosa para difundir información entre sus millones de usuarios. Las herramientas de que surte para facilitarla son los *retweet* (RT), los *hashtag* (#) y derivados de estos, los *Trending*

Topics o temas de máxima actualidad y actividad en la red. En un análisis realizado sobre la transmisión de mensajes en este canal, se manifestó que se cumple escasamente la “ley de potencias” en la distribución de seguidores, que existe un grado bajo de reciprocidad en las relaciones, que la mayoría de los temas se corresponden con titulares de noticia o similares, que el concepto de popularidad se obtiene tanto por número de seguidores como por la popularidad de un *tweet*, y que tras el primer *retweet*, un mensaje es re-difundido por otros a sus seguidores instantáneamente. En Twitter aunque los usuarios con más seguidores tienen más posibilidad de obtener *retweets* -redifusión- de sus mensajes (por el número de seguidores en sí, obviamente), algunos individuos son capaces de conseguir muchos *retweets* con un bajo número de seguidores debido probablemente a que su contenido se considera interesante y contagioso.

Es imprescindible remarcar que, en lo que respecta al comportamiento de las Redes Sociales en Internet comparado con el comportamiento de las redes sociales físicas, las diferencias en las conclusiones de diversos estudios dedicados al estudio de dicho comportamiento, difieren en sus conclusiones, y a más, no se encuentra siempre paridad en los resultados concluyentes entre los mismas (la mayor digresión se halla entorno al concepto de la Ley de Potencias, que de existir, pudiera no ser la única condición significativa ni decisoria en aras de conseguir que una idea cualquiera se propague por la red). No obstante, si coinciden en aspectos que son sin duda del mayor interés:

- Más que el/los usuario/s influyente/s se trata de un número considerable de nodos que inician un proceso de difusión que concluye en una cascada global.
- La popularidad o número de seguidores así como el volumen de actividad de un usuario, no es sinónimo de influencia.
- La naturaleza del contenido es básica para su difusión, siendo el de carácter noticiable el más propenso a conseguirlo.
- Los influyentes no siempre consiguen que su contenido sea retwiteado. Por el contrario, usuarios anónimos alcanzan un

Trending Topic, el cual es resultado de que varios usuarios lo propaguen a la vez.

- Los lazos débiles y los *early adopter* son de vital importancia en el proceso de difusión de innovaciones.
- La teoría “*two-step-flow*” parece quedar relegada en estas redes, en las que los “influyentes” no son únicamente los responsables de la difusión desde los medios, sino que, una vez más, los escasos vínculos (lazos débiles) posibilitan las interacciones entre comunidades (individuos densamente conectados).

Hemos visto como en el análisis de Redes Sociales como método de predicción electoral ha obtenido resultados dispares. Mientras que siguiendo los métodos tradicionales de encuesta, un deseable margen de error oscila entre el 2 o 3%, en este tipo de análisis el margen de error puede llegar a alcanzar el 10%. No obstante, también se ha constatado que las encuestas de opinión tradicionales no son fiables al 100% ni en todos casos, siendo el ejemplo español más reciente el error de predicción que mostrarón todas las encuestas realizadas por distintas agencias con motivo de las Elecciones al Parlamento de Andalucía 2012.

En lo que se refiere al análisis de redes como predicción electoral, centrándonos en la red social Twitter, y a la luz de la disparidad encontrada en estos hallazgos, coincidentes como en el caso de las Elecciones Alemanas de 2009 y desviados como en el caso de varias Elecciones en EEUU, podríamos apuntar que existen corrientes opuestas (o cuanto menos, dispares) en este campo, al igual que sucedía en su momento con la convivencia de las corrientes ciberoptimista, ciberpesimista y escéptica, en el estudio de la Ciberdemocracia:

- Los que afirman que el análisis de Redes Sociales constituye una herramienta rápida, barata y fiable de predicción de la opinión pública y a más, de predicción electoral.

- Los que atestiguan que dicho análisis, aunque posible, dados los métodos que hasta ahora se han utilizado, no puede tomarse como legítimo al superar su margen de error valores difícilmente admitidos.

Lo que podemos argumentar según los autores consultados es que no siempre es posible predecir elecciones a partir de los datos extraídos de la participación de los usuarios en las redes sociales. Por una parte, los resultados obtenidos en un ámbito de observación, no son directamente extrapolables a otro ni servirían por esta razón los mismos métodos, tanto porque sus sistemas electorales son heterogéneos como porque los espacios físicos en los que se producen presentan características demográficas desemejantes. Por otra, los métodos de correlación que se emplean son también diferentes, siendo en unos casos los datos obtenidos comparados con los propios resultados electorales definitivos y en otros con las encuestas preelectorales. Y por último, las herramientas y métodos de análisis, novedosos hasta el momento, deben ser testados y continuamente mejorados para asegurar el veraz resultado de las informaciones desprendidas. Además, hemos de tener en cuenta que no todas las franjas de edad son usuarios activos en estas redes sociales, apareciendo los mayores de 50 años como los menos propensos a participar y compartir así sus opiniones y tendencias de voto, y no encontrándose representados por tanto en estos canales de difusión.

En nuestro caso particular, se han monitorizado cuatro procesos electorales en España, celebrados en 2010 y 2011, siendo uno de ellos de ámbito local y el resto de ámbito regional: Elecciones Generales 2011, Elecciones Autonómicas a la Comunidad de Madrid 2011, Elecciones Autonómicas a la Región de Murcia 2011 y Elecciones al Parlamento de Cataluña 2010.

Fijándonos en métodos utilizados en varios de los trabajos mencionados a lo largo de esta obra, se decidió seguir como base inicial la técnica empleada en el publicado por Tumasjan et al., (2010) en el que sus autores concluyen que el número de *tweets* en los que se menciona a uno u otra opción política puede considerarse como un reflejo válido y previsible de la intención de voto, y por ende, del resultado

electoral, acercándose a los métodos predictivos obtenidos por las encuestas tradicionales. En dichas elecciones se toman como referencia únicamente las menciones a los partidos, para la investigación desarrollada se han tomado en cuenta la presencia y difusión tanto de partidos como de candidatos. Además, dichas menciones se han contabilizado de forma única y en análisis paralelos, junto con el resto de actividad Twitter, como Followers, RT's, tweets publicados y Reply's. Cada una de ellas se ha definido como variable independiente.

La captura de los *tweets* se ha realizado empleando el API de Twitter. Se han descargado con total continuidad, disponiendo de servidores y otros medios técnicos destinados a evitar rupturas o paradas del servicio que hubieran provocado, en caso de producirse, pérdidas de información.

El periodo de captura para todos los casos analizados se ha establecido en los quince días de campaña previos al día de celebración de cada una de los procesos electorales monitorizados. Además, todas las medidas se realizaron por días para que puedan ser agrupadas de una forma más flexible y en caso de necesidad, analizadas individualmente. De hecho, se llevo a cabo un análisis paralelo en el que el periodo de captura se redujo a la mitad, siendo contabilizados los datos referidos a la semana previa a la jornada electoral.

La *query* lanzada para la búsqueda contiene los nombres de los perfiles de candidatos y partidos (en sus tres posibles variantes: texto, @ y #), y los *hashtags* de la campaña. Con el fin de minimizar la duplicación de perfiles o el seguimiento de perfiles falsos, los datos se han obtenido de las páginas Web's oficiales de cada opción política. No obstante, en algunos casos se han incorporado *hashtags* creados por los propios usuarios y que se han considerado por su aceptación en la red y, por tanto, por su capacidad de difusión.

Las variables independientes clasificadas para cada opción política han sido las siguientes: Followers candidato, Followers partido, Tweets publicados por candidato, Tweets publicado por partido, Menciones a candidato (en @, en # y en

texto), Menciones a partido (en @, en # y en texto), Menciones a hashtags de campaña, RT a candidatos, RT a partidos, Reply a candidatos y Reply a partidos.

Se ha diseñado un método simultáneo comparativo cuyo objetivo ha sido tratar de conseguir un peso específico para cada una de las variables independientes en el conjunto de la estrategia de campaña articulada en Twitter, determinando así el orden de relevancia que supuestamente éstas han ocupado en la misma. Para ello, se ha asignado a cada una de las variables un valor de ponderación, en una escala de 1 a 10.

Valor de Ponderación asignada a cada variable	
Ponderación	Variable
1	Followers candidato
1	Followers partido
1	Tweets publicados por candidato
1	Tweets publicado por partido
3	Menciones a candidato (en @, en # y en texto)
5	Menciones a partido (en @, en # y en texto)
9	Menciones a hashtags de campaña
4	RT a candidatos
6	RT a partidos
2	Reply a candidatos
4	Reply a partidos

Si bien podría argumentarse que los valores de ponderación incluyen un grado de subjetividad de aquel que los asigna, estos se han establecido teniendo en cuenta conclusiones de estudios publicados por diferentes autores referenciados en el marco teórico de esta investigación.

Se ha desestimado el “análisis sentimental” o *sentiment analysis* por varias razones. La primera y no desdeñable, es la complejidad de su estudio, y la segunda y no menos importante es la constatación de que este tipo de análisis no incorpora mejoras en comparación con el análisis único de los tweets.

En términos generales, sin duda los análisis para Elecciones Generales 2011 son los que han obtenido mejores resultados, encontrándose los mismos cercanos al resultado electoral definitivo y siendo las diferencias encontradas similares y, en varios casos, menores que las obtenidas por las encuestas de opinión. Dicha característica se observa también en las Elecciones al Parlamento de Cataluña 2010. En el resto de procesos, Elecciones Autonómicas a la Comunidad de Madrid 2011 y Elecciones Autonómicas a la Región de Murcia 2011, los resultados de la actividad Twitter presentan grandes diferencias para varias de sus opciones políticas.

El MAE obtenido en cada una de ellas es el siguiente:

- Elecciones Generales 2011 – 1,57%
- Elecciones Autonómicas a la Comunidad de Madrid 2011 – 4,10%
- Elecciones Autonómicas a la Región de Murcia 2011 – 4,63%
- Elecciones al Parlamento de Cataluña 2010 – 2,98%

Las candidaturas que presentan mayores desviaciones en todos los casos son las correspondientes a IU (Izquierda Unida) y UPYD (Unión Progreso y Democracia). En el primer caso, la tendencia es a la infravaloración y en el segundo a la sobrevaloración. Entre las posibles causas podría argumentarse que IU es un partido político que mantiene un voto cautivo cuya mayoría no está representada en esta Red al no participar masiva o proporcionalmente en ella. UPYD, por el contrario, es un partido joven, de reciente aparición, con un electorado joven y participe en Twitter. Está demostrado además que los partidos muestran una mayor tendencia a hacer uso de Internet, dada la escasa capacidad que tienen de ser amplificados por los medios de comunicación tradicionales como la prensa escrita y la televisión.

Es preciso señalar las particularidades intrínsecas a cada elección, así como aquellas que han afectado a las particularidades de su monitorización en Twitter, a saber:

- En las Elecciones Generales 2011, el tamaño de la muestra supera al resto de las elecciones estudiadas, tanto en porcentaje de población usuario de la Red, como en volumen de tweets y usuarios únicos.
- En las Elecciones al Parlamento de Cataluña 2010, el número de partidos u opciones políticas monitorizadas prácticamente dobla el número de cualquiera del resto de procesos analizados. Dados los buenos resultados obtenidos en esta elección, esta realidad podría en parte contradecir el trabajo de Jungherr et al. (2011) cuando afirman que la inclusión de nuevos partidos, como el Partido Pirata, puede modificar considerablemente las predicciones basadas en el análisis Twitter. Es preciso de nuevo recordar que se ha de tener en cuenta el peso que los partidos minoritarios alcanzan en Twitter, al ser una de sus plataformas Red accesibles para la difusión de sus propuestas, y que por tanto, su presencia en ella se encontraría sobredimensionada respecto su peso real. Además, en esta ocasión sólo se encuentra monitorizada la actividad Twitter referida a Partido.
- En las Elecciones Autonómicas a la Región de Murcia 2010, la opción política ganadora, el Partido Popular, no mantuvo presencia alguna en esta red (ni el partido ni su candidato).

Las conclusiones generales obtenidas a tenor de la observación comparada de lo que ha sucedido en los cuatro procesos electorales estudiados y organizadas entorno a los distintos análisis desempeñados, son las siguientes:

- Resultado Twitter (Candidato+Partido)

- Cuando se aplica la ponderación, las diferencias mejoran en todos los casos, y en aquellos en los que empeora, lo hace ligeramente con un $\pm 0,50\%$ ²⁴² (PP Madrid e IUV-RM).
- La opción política que mantiene mejores y constantes resultados en todas las elecciones, es el PSOE (Partido Socialista Obrero Español).
- Cuando el universo de la muestra disminuye, referido este al porcentaje de población que expresa pertenecer a la Red Twitter, las expectativas Twitter aunque empeoran en comparación con universos mayores, no lo hacen de forma progresiva, como puede deducirse del hecho de que los resultados de la actividad Twitter son mejores en las Elecciones al Parlamento de Cataluña 2010 que en las Elecciones Autonómicas a la Comunidad de Madrid 2011, siendo la población representada en Twitter mayor en esta última que en la anterior. Además, en las Elecciones Autonómicas a la Región de Murcia 2011, cuya población en Twitter es muchísimo menor que en las de la Comunidad de Madrid, los resultados no empeoran numéricamente en el mismo orden o porcentaje en que dicha población decrece (es preciso recordar aquí, que en la Región de Murcia tanto el partido como el candidato ganador y que se presentaba a su reelección, carecían de presencia en esta red).
- Resultado Twitter (Candidato+Partido) sólo Menciones.
 - Cuando se toman en cuenta sólo las menciones, sólo se observa mejoría para la opción PP Madrid.
- Resultado Twitter Candidato y Resultado Twitter Partido. Cuando se toman en cuenta las actividades sólo por Partido o sólo por Candidato, la

²⁴² Tomando como referencia los resultados sin ponderación

tendencia general es a empeorar en todas las elecciones y para todas las opciones políticas, con las siguientes excepciones:

- PP de Madrid mejora para Candidato y empeora notablemente para Partido.
- IU (Izquierda Unida) mantiene valores similares en el caso de Partido en todas las elecciones analizadas
- UPYD mantiene siempre valores similares para Partido.

También se ha constatado que en aquellas opciones políticas en las que la actividad Twitter general ha sido muy reducida en comparación con su masa de posibles votantes, las diferencias comparativas con los resultados electorales son notables, siendo éstas siempre en carácter negativo u opción política infradimensionada (IU). De igual forma pero en sentido contrario, en aquellas opciones políticas en las que la actividad ha sido incrementalmente desproporcionada las diferencias, también notables, con los resultados electorales adquieren un carácter positivo u opción política sobredimensionada (UPYD).

Al reducir a la mitad el periodo establecido para la captura de datos, tomando los datos de la semana previa a la jornada electoral, se ha comprobado que los resultados mejoran en procesos de volumen de actividad considerable, como es el caso de las Elecciones Generales 2011 y se mantienen o empeoran (dado que las desviaciones continúan en uno u otro sentido) en elecciones de menor envergadura o de ámbito regional.

Más allá de las variables y opciones políticas vigiladas, se han encontrado indicios reveladores acerca de las posibles causas de la eficacia o desviación en el análisis de Twitter como herramienta de predicción, referidas al volumen de datos manejados en relación con el volumen de la muestra. Los hallazgos han sido los siguientes:

- El volumen de tweets recolectados es de suma relevancia para la lograr un buen resultado, coincidiendo con las conclusiones del

trabajo elaborado por Twitminster (2010). Así, la precisión es muy buena en Elecciones Generales 2011, en la que se supera el millón de tweets tratados. Para ámbitos regionales, en los que dicho volumen disminuye considerablemente, no existe relación alguna entre el volumen de datos y las poblaciones de estudio.

- El número de partidos que se incluyen en el estudio puede afectar a los resultados. La introducción de partidos minoritarios o con escasa o nula representación parlamentaria, así como de partidos de reciente aparición, que tuvieron cierta presencia o viralidad en las redes sociales, puede modificar substancialmente el análisis en Twitter. En esta investigación, nos estaríamos refiriendo a Equo (Elecciones Generales 2011), SI (Elecciones al Parlamento de Cataluña) e UPYD (Elecciones Autonómicas a la Región de Murcia 2011).
- Se ha hallado una causalidad directa entre la eficacia del análisis electoral en Twitter y la relación entre “usuarios únicos” y “usuarios de Twitter” y “volumen de tweets”, de forma que cuanto más alejados de cero se encuentren los valores de estas relaciones menor es la eficacia de su predicción, y viceversa. Dicha condición se cumple en las cuatro elecciones monitorizadas.

Son sin duda estos últimos hallazgos anotados los que se consideran de mayor relevancia y concluyentes en la investigación, dejando constancia de que existen tres factores determinantes en el éxito del uso de Twitter como herramienta de predicción electoral: el volumen de tweets capturados, el número de opciones políticas monitorizadas y la relación “Usuarios únicos/Volumen de tweets”. A éstos, podríamos añadir el periodo de captura de datos, su amplitud o extensión temporal, y su situación de cercanía o lejanía de la jornada electoral.

Se hace necesario dejar constancia de que este método de análisis basado en el cómputo de menciones que recibe una determinada opción política cuenta con la limitación o barrera impuesta por el desconocimiento del sesgo político de los

participantes en esta red y por el desconocimiento del peso que los seguidores (muy activos en esta red) de los partidos y/o candidatos adquieren en el volumen de tweets, esto es, cuántos de los tweets publicados son de su propiedad.

Dadas las divergencias encontradas en la presente investigación junto con las también dispares anunciadas en previos trabajos de varios autores vigilando procesos electorales distintos, se hace necesaria la labor de nuevos investigadores que continúen perfeccionando las metodologías iniciadas incorporando sus propios métodos y aportando nuevas y decisivas conclusiones. Por otra parte, sería de gran interés en este campo dedicado a la predicción mediante análisis de redes, la introducción paralela de los datos conseguidos a través de la vigilancia de otras redes sociales, como Facebook, Twitter y Google Trends, tratando de encontrar similitudes y divergencias entre sus resultados, siendo un mismo proceso electoral, el ámbito de estudio para todas ellas.

Dicha futura línea de investigación tendría que ser capaz de analizar y cuantificar aspectos muy concretos en lo que se refiere al modo de difusión de un mensaje político, la forma de su transmisión, su alcance o longitud de la cadena y el papel que los diferentes tipos de usuarios desempeñan.

Así, una propuesta de mejora en la investigación podría fijarse en los siguientes puntos:

- Definir y delimitar la idea que se pretende transmitir por la red. Por ejemplo, los mensajes lanzados por un partido político en campaña.
- Anclar la red, es decir, tomar a un actor influyente en la misma -que no pertenezca al equipo de campaña pero que se tome por influyente en ese tema- y seguir el flujo de la difusión para analizar su patrón de conducta.
- Cuantificar la longitud alcanzada por la cadena de difusión, los tipos de vínculos que forman parte en ella y el tiempo de reacción en cada caso -mensaje-.

- Analizar las reacciones y cuantificarlas. Se trata de realizar un análisis semántico que nos informe sobre si las reacciones - comentarios- en cada punto de la cadena han sido positivos, negativos o neutros.
- Categorizar a los participantes en dicha red, definiendo su sesgo político y su volumen de actividad.

Por todo lo dicho, se abre un campo de especial interés y relevancia de cara al futuro, en el que tal vez llegue a ser posible sustituir o convivir con las encuestas preelectorales, que hasta hoy han sido una fuente fiable de la tendencia electoral. Campo en el que esta investigación espera haber contribuido con sus resultados, en espera de un desarrollo de metodologías de análisis más sofisticadas.

Bibliografía

7. Bibliografía

7.1. Obras

1. ANDUIZA, Eva, CRESPO, Ismael y MÉNDEZ, Mónica. (2009). *Metodología de la Ciencia Política*. CIS, Centro de Investigaciones Sociológicas. En Cuadernos Metodológicos, 2ª edición revisada. Madrid
2. CALDERÓN, César, LORENZO, Sebastián, y 15 autores más (2010) *Open Government-Gobierno Abierto*. Editorial Algón Editores. Madrid
3. CHISTAKIS, Nicholas A. y FOWLER, James H. (2010) *Conectados, el sorprendente poder de las redes sociales y cómo nos afectan*. Editorial Taurus. Madrid
4. FERNÁNDEZ, Montse y CONGOSTO, Mariluz (2012) “Twitter en las elecciones autonómicas 2011 en la Comunidad de Madrid”. *Las campañas electorales y sus efectos sobre el voto en la Comunidad de Madrid*. Editorial Fragua. Madrid
5. CRESPO, Ismael, et al. (2002) *Las Campañas Electorales y sus Efectos en la Decisión del Voto Vol. II*. Editorial Tirant lo Blanch. Valencia
6. GLADWELL, Malcom. (2002) *The Tipping Point: How Little Things Can Make a Big Difference*. (Edición en castellano “La clave del éxito”. 2007. Editorial Taurus)
7. KELLER, Ed y BERRY, Jonh. (2003) *The influentials. One American in ten tells his other nine how to vote, where to eat and what to buy*. The Free Press. Detroit
8. ROGERS, Everett M. *The diffusions of innovations*. Fourth Edition. The Free Press. Detroit
9. TOHARIA, Juan M. (2012). *Pulso de España (enero 2011-marzo2012)*. Biblioteca Nueva. Madrid.
10. ZAPATA, Oscar A. (2006). *La aventura del pensamiento crítico: herramientas para elaborar tesis e investigaciones socioeducativas*. Editorial Pax México. México

7.2. Papers

1. ADAMIC, Lada y GLANCE, Natalie. (2005). "The political blogosphere and the 2004 U.S. election: Divided they blog". *In Proc. 3rd Intl. Workshop on Link Discovery (LinkKDD)*, 36–43.
2. ALSTYNE, Van y BRYNJOLFSSON, Eric. 2004. "Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities". *Management Science*, *Forthcoming*. [Internet]. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=756445 (Consultado el 21 de febrero de 2010)
3. ASUR, Sitaram; HUBERMAN, Bernardo A.; SZABÓ, Gábor y WANG, Chunyan (2011) "Trends in Social Media Persistence and Decay". *Presentado en CoRR, 2011*.
4. ASUR, Sitaram y HUBERMAN, Bernardo A. (2010) "Predicting the Future with Social Media". *Technical report. CoRR abs/1003.5699*. [Internet]. <http://arxiv.org/abs/1003.5699v1> (Consultado el 18 de mayo de 2011)
5. BACKSTROM, Lars, Eytan BAKSHY, Jon KLEINBERG, Thomas M. LENTO e Itamar ROSENN (2011). "Center of Attention: How Facebook Users Allocate Attention across Friends". *Association for the Advancement of Artificial Intelligence (www.aaai.org)*
6. BARBERÁ, Pablo y RIVERO, Gonzalo (2012). "Un tweet un voto. Desigualdad en la discusión política en Twitter". Ponencia presentada en el *I Congreso Internacional en Comunicación Política y Estrategias de Campaña*, ALICE.
7. BARNES, J. A. "Class and Committees in a Norwegian Island Parish". *Human Relations*, Vol 7, pp. 39-58
8. BOLLEN, Johan; MAO, Huina y ZENG, Xiao-Jun (2010). "Twitter mood predicts the stock market". *Journal of Computational Science*, 2(1), March 2011, pp. 1-8
9. BOND, Robert M.; FARISS, Christopher J.; JONES, Jason J.; KRAMER, AdamD. I.; MARLOW, Cameron; SETTLE, Jaime E. y FOWLER, James H.

- (2012) "A 61-million-person experiment in social influence and political mobilization". *Nature*, 13 de septiembre de 2012, Vol 489, 295-298
10. CHA, Meeyoung; HADDADI, Hamed; BENEVENUTO, Fabrício y GUMMADI Kristina P. (2010) "Measuring user influence in Twitter: The million followers fallacy". *Association for the Advancement of Artificial Intelligence*.
 11. CHA, Meeyoung; MISLOVE, Alan; ADAMS, Ben, y GUMMADI, Krishna P. (2008). "Characterizing Social Cascades in Flickr" *WOSN'08*, August 18, 2008, Seattle, Washington, USA. Copyright 2008 ACM 978-1-60558-182-8/08/08
 12. CHANG, Jonathan; ROSENN, Itamar; BACKSTROM, Lars; y MARLOW, Cameron (2010). "epluribus: Ethnicity on social networks". In *Proc. 4th International Conference on Weblogs and Social Media*.
 13. CHOI, Hyunyoung y VARIAN, Hall (2009) "Predicting the Present with Google Trends". In *Google Research Blog*, 2 de abril de 2009. [Internet]. http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en//googleblogs/pdfs/google_predicting_the_present.pdf (Consultado el 15 de enero de 2011).
 14. COLBAUGH, Richard y GLASS, Kristin (2009) "Predictive Analysis for Social Diffusion: The Role of Network Communities" *eprint arXiv:0912.5242*. [Internet]. <http://arxiv.org/abs/0912.5242>. (Consultado el 12 de diciembre de 2009)
 15. CONGOSTO, Mariluz y ARAGÓN, Pablo (2012) "Twitter, del sondeo a la sonda: nuevos canales de opinión, nuevos métodos de análisis". Conferencia presentada en el *I Congreso Internacional en Comunicación Política y Estrategias de Campaña*, Madrid.
 16. CONOVER, M.; RATKIEWICZ, J.; FRANCISCO, M.; GONÇALVES, B.; Flammini, A. y MENCZER, F. (2011) "Political polarization on twitter". *5th International Conference on Weblogs and Social Media (ICWSM)*. Barcelona, España
 17. DADER, José Luis (2001). "La Ciberdemocracia posible: Reflexión prospectiva a partir de la experiencia en España". *Cuadernos de Información y Comunicación (CIC)* Núm. 6, pp.177- 220.

18. DESART, Jay y HOLBROOK, Thomas (2003). "Campaigns, Polls, and the States: Assessing the Accuracy of Statewide Presidential 'Trial-Heat' Polls." *Political Research Quarterly*, Vol. 56, No. 4 (Dec., 2003), pp. 431-439. Publicado por Sage Publications, Inc. en nombre de The University of UtahStable.
19. FREEMAN, Linton C. (n.d.) "El problema de la centralidad". *Social Networks, 1 (1978/79) 215-239 Departamento de Relaciones Sociales, Price Hall, Building # 40, Bethlehem, P.A. 18015. USA.* (Traducido al castellano por Reyes Herrero)
20. GAYO-AVELLO, Daniel; METAXAS, Panagiotis T.; y MUSTAFARAJ, Eni. (2011) "Limits of electoral predictions using Twitter". *Association for the Advancement of Artificial Intelligence*
21. GAYO-AVELLO, Daniel (2011) "A warning against converting Social Media into the next Literary Digest". *Communications of the ACM*
22. GIBSON, Rachel y MCCALLISTER, Ian (2005) "Does Cyber-Campaigning Win Votes?" *Online Communication in the 2004 Australian Election.2005.* Research School of Social Sciences Australian National University. Canberra. ACT 0200. Australia
23. GINSBERG, Jeremy; MOHEBBI, Matthew H.; PATEL, Rajan S.; BRAMMER, Lynnette; SMOLINSKI, Mark S. y BRILLIANT, Larry (2009). "Detecting influenza epidemics using search engine query data". *Nature*, Vol. 457, 19 de febrero de 2009
24. GOLBECK, Jen y HANSEN, Derek L. (2011) "Computing political preference among twitter followers". *In Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*. ACM, New York, NY, USA, 1105-1108. DOI=10.1145/1978942.1979106
25. GONÇALVES, Bruno; PERRA, Nicola y VESPIGNANI, Alessandro (2011) "Modeling Users' Activity on Twitter Networks: Validation of Dunbar's Number". *PLoS ONE* 6(8): e22656. doi:10.1371/journal.pone.0022656

26. GRANOVETTER, Mark S. (1973). "The strength of weak ties". *American Journal of Sociology*; Vol 78, nº 6. (pp. 1360 - 1380). (Traducido al castellano por Mª Ángeles García Verdasco)
27. HEYLIGHEN, Francis (1999). "What makes a meme successful? Selection criteria for cultural evolution". *Proc. 15th Int. Congress on Cybernetics Association Internat. de Cybernétique*, Namur, 1999, p. 418- 423
28. JARIEGO, Isidro M., "Internet, amigos y bacterias: la alargada sombra de Stanley Milgram". *Araucaria*, Segundo semestre, año/ vol 4, número 010, Universidad de Sevilla, Sevilla, España
29. KATZ, Elihu y LAZARSFELD, Paul F. (1955), "Personal Influence; the Part Played by People in the Flow of Mass Communications". *Glencoe, IL: Free Press*
30. KWAK, Haewoon; LEE, Changhyun; PARK, Hosung y MOON Sue. (2010) "What is Twitter, a Social Network or a News Media?" *International World Wide Web Conference Committee. (IW3C2)*. WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA. ACM 978-1-60558-799-8/10/04.
31. KAVANAGH, Dennis (1994). "Las encuestas de opinión pública". *Estudios públicos*, 53, pp.145-169. [Internet]. <http://bit.ly/HvGgs8>. (Consultado el 2 de abril de 2011)
32. KREBS, Valdis. (2005). "The Link between Social Interaction and Political Choice". *Extreme Democracy, The Book and Discussion Forum for Networked Activists*. Editado por Jon Lebkowsky y Mr. Mitch Ratcliffe - Capítulo 9. Internet/Media Strategies, Inc. [Internet]. <http://www.extremedemocracy.com/chapters/Chapter%20Nine-Krebs.pdf> (Consultado el 18 de mayo de 2010)
33. LAZER, David; RUBINEAU, Brian; CHETKOVICH, Carol; KATZ, Nancy y NEBLO, Michael (2009) "The coevolution of networks and political attitudes". *Southern Illinois University Carbondale. Working Papers. Posted at OpenSIUC*.
34. LAU, Richard R. (1994). "An Analysis of the Accuracy of 'Trial-Heat' Polls during the 1992 Presidential Election." *Public Opinion Quarterly*, 58, pp. 2-20.

35. LUI, Catherine; METAXAS, Panagiotis T.; y MUSTAFARAJ, Eni (2011) “On the predictability of the U.S. elections through search volume activity”. *e-Society Conference*.
36. MANTILLA, Lya Fernández y PINILLA, Katherine Flórez. (2008). “¿Qué evalúa el ciudadano al momento de votar? Algunas apreciaciones desde el Enfoque Racional”. *Universidad Autónoma de Bucaramanga*. [Internet]. <http://caribdis.unab.edu.co/pls/portal/docs/PAGE/REFLEXIONPOLITICA/ANTERIORES/REVISTA1019/QUE%20EVALUA%20EL%20CIUDADANO%20EN%20EL%20MOMENTO%20DE%20VOTAR.PDF>. (Consultado el 15 de junio de 2010)
37. JUNGHERR, Andreas; J’URGENSE, Pascal y SCHOEN, Harald (2011). “Why the Pirate Party won the german election of 2009 or the trouble with predictions: A response to Tumasjan, a., Sprenger, t. o., Sander, p. g., & Welpe, i. m. ’Predicting elections with twitter: What 140 characters reveal about political sentiment’”. *Social Science Computer Review*.
38. MILGRAM, Stanley (1967). “El problema del mundo pequeño”. *Reproducido con permiso de Phychology Today Magazine. Copyright 1967, Sussex Publishers Inc. Traducción de Isidro Maya Jariego. Araucaria*, Segundo semestre, año/vol 4, número 010. Universidad de Sevilla. Pp. 15-28
39. MUHLBERGER, Peter (2004). “Polarization of Political Attitudes and Values on the Internet”. *All academic research*. Paper to be presented at the 2004 International Communication Association Annual Meeting in New Orleans, Louisiana [Internet]. http://www.allacademic.com/meta/p_mla_apa_research_citation/1/1/2/5/2/pages112523/p112523-27.php (Consultado el 14 de febrero de 2010)
40. NEWMAN, M. E. J. y PARK, Juyong. (2003). “Why social networks are different from other types of networks”. *Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109 and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501*. [Internet]. http://arxiv.org/PS_cache/cond-mat/pdf/0305/0305612v1.pdf (Consultado el 22 de mayo de 2010)

41. NORRIS, Pippa (2003) “Preaching to the converted?” *Party Politics* VOL 9. N° 1, pp.21–45. SAGE Publications
42. OLSON, Steve y BUNNETT, Will (2010) “Facebook predicts electoral victory or not”. *Trilogy*. [Internet]. http://www.trilogyinteractive.com/assets/pdf/facebookpredictselectoralvictory_final.pdf (Consultado el 3 de agosto de 2011)
43. O’CONNOR, Brendan; RAMMATH, Balasubramanyan; BRYEN, R. Routledge y NOAH, A. Smith (2010) “From tweets to polls: Linking text sentiment to public opinion time series”. *Proc. of 4th ICWSM, 122–129. AAAI Press*
44. ROMERO, Daniel M.; GALUBA, Wojciech; ASUR, Sitaram y HUBERMAN, Bernardo A. (2011) “Influence and Passivity in Social Media”. *20th International World Wide Web Conference (WWW’11), 2011*.
45. RODRÍGUEZ, Jorge (2004). “Una web muy fragmentada”. *Cátedra de procesamiento de datos*. [Internet]. <http://www.ilhn.com/datos/teoricos/archives/001967.php> (Consultado el 19 de julio de 2010)
46. RHODE, Paul W. y STRUMPF, Koleman S. (2004) “Historical Presidential Betting Markets”. [Internet]. http://www.unc.edu/~cigar/papers/BettingPaper_final%28JEP_Resubmit%29.pdf (Consultado el 7 de marzo de 2012)
47. ROTHCHILD, David (2009) “Forecasting Elections. Comparing Predictions Markets, Polls and their biases”. *Public Opinion Quarterly*, Vol. 73, No. 5, 2009, pp.895–916. Publicado por Oxford University Press en nombre de American Association for Public Opinion Research
48. SUN, Eric; ROSENN, Itamar; MARLOW, Cameron y LENTO, Thomas M (2009). “Gesundheit! Modeling Contagion through Facebook News Feed”. Department of Statistics, Stanford University. *Association for the Advancement of Artificial Intelligence*.
49. SANTOS, Félix R. (1989) “El concepto de Red Social”. *Reis: Revista española de investigaciones sociológicas*, ISSN 0210-5233, N° 48, 1989, págs. 137-152

50. SUNSTEIN, Cass R. (2000). “Deliberative Trouble? Why Groups Go to Extremes”. *Yale L.J.* 71. [Internet]. <http://www.yalelawjournal.org/the-yale-law-journal/content-pages/deliberative-trouble?-why-groups-go-to-extremes/> (Consultado el 12 de noviembre de 2011)
51. SUNSTEIN, Cass R. (2005) “La ley de polarización de grupos”. Traducido al castellano por Juan F. González Bertomeu. *Revista jurídica de la Universidad de Palermo*. pp. 55-70. [Internet]. http://www.palermo.edu/derecho/publicaciones/pdfs/revista_juridica/n6N1-Octubre2005/061Juridica03.pdf (Consultado el 2 de mayo de 2012)
52. TRUONG, Hung (2009). “Network Structure. Diffusion and Growth Over Time in a Facebook Application”. *Universidad de Michigan*. [Internet]. <http://www-personal.umich.edu/~ladamic/courses/networks/si508f07/projects/facebook.pdf>
53. TUMASJAN, Andranik; SPRENGER, Tim O., SANDNER, Philipp G. y WELPE, Isabell M. (2010) “Predicting elections with twitter: What 140 characters reveal about political sentiment”. *Proc. of 4th ICWSM*, 178–185. AAAI Press.
54. TWEETMINSTER (2010) “Is word-of-mouth correlated to General Election results? The results are in.” *Publicado por Tweetminster en Scribd*. [Internet]. <http://es.scribd.com/doc/31208748/Tweetminster-Predicts-Findings> (Consultado el 7 de agosto de 2011)
55. UGANDER, Johan; BACKSTROM, Lars; MARLOW, Cameron y KLEINBERG, Jon (2012). “Structural diversity in social contagion”. Center for Applied Mathematics and Department of Computer Science, Cornell University, Ithaca, NY 14853; and Facebook, Menlo Park, CA 94025. *Edited by Ronald L. Graham, University of California at San Diego, La Jolla, CA*,
56. VILPPONEN, Antti; WINTER, Susanna y SUNDQUVIST, Sanna (2006). “Electronic Word-of-Mouth in Online Environments: Exploring Referral Network Structure and Adoption Behavior”. *Journal of Interactive Advertising*, Vol. 6, N° 2 (Spring 2006), pp.63-77.

57. WATTS, Duncan J. y DODDS, Petter S. (2007). "Influentials, Networks, and Public Opinion Formation". *Journal of Consumer Research, Inc.* Vol. 34. pp. 441-458, Diciembre de 2007
58. WILLIAMS, Christine B. y GULATI, Girish "Jeff". (2009). "Explaining Facebook Support in the 2008 Congressional Election Cycle". *Working Papers. Paper 26. Southern Illinois University. Carbondale.* [Internet]. http://opensiuc.lib.siu.edu/pn_wp/26 (Consultado el 19 de septiembre de 2010)
59. WILLIAMS, Christine B. y GULATI, Girish "Jeff" (2008). "What is a Social Network Worth? Facebook and Vote Share in the 2008 Presidential Primaries". *Annual Meeting of the American Political Science Association, 1-17. Boston, MA*
60. YU, Bei; KAUFMAN, Stefan y DIERMEIER, Daniel. (2008) "Exploring the Characteristics of Opinion Expressions for Political Opinion Classification". *Digital Government Conference '08*, May 18–21, 2008, Montreal, Canada
61. ZHANG, Xue; FUEHRES, Hauke y GLOOR, Peter A. (2010). "Predicting Stock Market Indicators Through Twitter «I hope it is not as bad as I fear»" *Elsevier*. [Internet]. http://www.ickn.org/documents/COINs2010_Twitter4.pdf (Consultado el 9 de julio de 2011)
62. ZHU, Haiyi, HUBERMAN, Bernardo A. y LUON, Yarun. (2012) "To Switch or Not To Switch: Understanding Social Influence in Recommender Systems". *CHI 2012*, May 5-10, 2012, Austin, TX, USA.

7.3. Artículos y consultas Web

1. ANDERSON, Chris (2004). "The Long Tail". *Wired*. Octubre de 2012. <http://www.wired.com/wired/archive/12.10/tail.html> (Consultado el 29/04/2012)
2. ARRIZABALAGA, M. (2012) "¿Por qué todas las encuestas fallaron en Andalucía?" *Diario ABC*. 26/03/2012. (Consultado el 14 de marzo de 2012).
3. AVELLANEDO, Antonio (2012) "25-M: el día en que se equivocaron todos (menos uno)". *Diario Público*. 26/03/2012. <http://www.publico.es/espana/427160/25-m-el-dia-en-que-se-equivocaron-todos-menos-uno>. (Consultado el 12 de abril de 2012)

4. CALDERÓN, Sara I. (2010) "Facebook Is an Increasingly Important Part of Elections Around the World". *Inside Facebook*. 25/01/2010. <http://www.insidefacebook.com/2010/01/25/facebook-is-an-increasingly-important-part-of-elections-around-the-world/> (Consultado el 3 de marzo de 2012)
5. CAIVANO, Michele (2012). "[INFOGRAPHIC] Top 15 Facebook Countries in 2012". *Fortunecat.it*. Enero de 2012. <http://www.fortunecat.it/facebook-statistics/> (Consultado el 3 de mayo de 2012)
6. CARR, Austin (2010) "Facebook, Twitter Election Results Prove Remarkably Accurate". *Fast Company*. 03/11/2010 <http://www.fastcompany.com/1699853/facebook-twitter-election-results-prove-remarkably-accurate> (Consultado el 30 de noviembre de 2010)
7. CHISTAKIS, Nicholas A. y FOWLER, James H. "On Facebook a picture is worth a thousand friends". *Connected*. 01/10/2009. http://www.connectedthebook.com/pages/links/tastes_on_facebook.html (Consultado el 11 de enero de 2010)
8. FREIRE, Juan. 09/11/2008. "Influencia de Internet en las elecciones en EEUU". Internet. 09/11/2008. <http://nomada.blogs.com/jfreire/2008/11/influencia-de-internet-en-las-elecciones-en-eeuu-cuadernos-de-periodistas.html> (Consultado en marzo de 2009)
9. GAD3 (2011) "#20n: análisis electoral: ¿Nos podemos fiar de las encuestas electorales?". *El Confidencial*. 07/09/2011. <http://www.elconfidencial.com/archivos/ec/2011101623gad3encuestas.pdf> (Consultado el 2 de abril de 2012)
10. GOLDSTEIN, Patrick y RAINEY, James (2010). "The 2010 elections: Twitter isn't a very reliable prediction tool." *The Envelope*. 03/11/2010. http://latimesblogs.latimes.com/the_big_picture/2010/11/the-2010-midterms-twitter-effect-not-a-very-reliable-election-prediction-tool.html (Consultado el 10 de diciembre de 2010)

11. JUANAS, María (2011). “Usuarios de Facebook en España en Julio de 2011” *Cuentamelared.com*. 03/08/2011. <http://www.cuentamelared.com/usuarios-facebook-julio/> (Consultado el 5 de mayo de 2012)
12. MARONESE, Armando (2012). “Ley de Potencias”. *Redación Digital*. 24/02/2012. <http://redactores.zoomblog.com/archivo/2012/02/24/ley-de-potencias.html> (Consultado el 29 de abril de 2012)
13. OXFORD ANALYTICA (2010). “Friendships in the digital age”. *Forbes*. 17/01/2010. <http://www.forbes.com/2010/01/17/internet-facebook-social-networking-business-friends-oxford.html> (Consultado el 14 de marzo de 2010)
14. QUIROGA, Águeda (UPF), MARTI, Joel (UAB), JARIEGO, Isidro Maya (US) y MOLINA, José Luis (UAB). (2005). “Talleres de autoformación con programas informáticos de análisis de redes sociales. Internet”. *Revista Redes*. Septiembre de 2005. http://revista-redes.rediris.es/webredes/red_tematica/talleresars.pdf. (Consultado el 14 de septiembre de 2010).
15. ROTHCHILD, David (2012). “Hollande favored as next French President”. *The Signal (Yahoo)*. 19/04/2012. <http://news.yahoo.com/blogs/signal/hollande-favored-next-french-president-150709105.html#more-id> (Consultado el 3 de mayo de 2012)
16. U.S Politics on Facebook (2010) “Snapshot: The Day After Election Day”. *Facebook*. 03/11/2010. <http://www.facebook.com/notes/us-politics-on-facebook/snapshot-the-day-after-election-day/448930025881>. (Consultado el 20 de noviembre de 2010)
17. ZARRELLA, Dan (2009). “The science of retweets”. *Web DanZarrella The Social Media Scientist*. <http://danzarrella.com/science-of-retweets.pdf> (Consultado el 15 de junio de 2010)
18. ZARRELLA, Dan (2010) “New data: Can Twitter predict elections?” *Web DanZarrella The Social Media Scientist*. (n.d.) http://danzarrella.com/new-data-can-twitter-predict-elections.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Fe

[ed%3A+DanZarrella+%28Dan+Zarrella%27s+Blog%29](#). (Consultado el 20 de noviembre de 2010)

19. ZOOM LABS (2012) “Análisis de Facebook en España”. *Zoom Labs*. 06/02/2012. <http://blog.zoomlabs.es/2012/02/analisis-de-facebook-en-espana/> (Consultado el 6 de abril de 2012)

7.4. Informes

1. ADIGITAL (2010). “Estudio uso de Twitter en España”. Julio de 2010. http://www.adigital.org/emailing/2010/docs/adigital_Estudio_Uso_Twitter_enEspana_2010.pdf (Consultado el 4 de abril de 2012)
2. AFHA (n.d.). “Metodología de Muestreo”. <http://www.cesdonbosco.com/cii/documentos/IT%20METODOLOG%CDA%20DE%20MUESTREO.pdf> (Consultado el 8 de enero de 2012)
3. ELOGIA e IAB (2011). “III Estudio sobre Redes Sociales e Internet”. 30/11/2011. <http://www.slideshare.net/elogia/informe-redes-sociales-iab-elogia-2011?ref=http://www.elogia.net/blog/estudio-habitos-redes-sociales-iab-2011/> (Consultado el 15 de abril de 2012)
4. FUNDACIÓN ORANGE (2011). “Informe anual sobre el desarrollo de la sociedad de la información en España 2011” <http://www.informeeespana.es/docs/eE2011.pdf> (Consultado el 14 de abril de 2012)
5. HAMPTON, Keith N., SESSIONS, Lauren F., HER, Eun J. y RAINIE, Lee. (2009). “Social Isolation and New Technology. How the internet and mobile phones impact Americans’ social networks”. *Pew Internet & American Life Project*. 04/11/2009. <http://www.pewinternet.org/Reports/2009/18--Social-Isolation-and-New-Technology.aspx> (Consultado el 3 de diciembre de 2009)
6. INE (Instituto Nacional de Estadística) (2011). “Encuesta sobre Equipamiento y Uso de Tecnologías de la Información y Comunicación en los hogares 2011”. Nota de prensa. [online] <http://www.ine.es/prensa/np678.pdf> (Consultado el 15 de abril de 2012)

7. MCCANN, Universal (2008) “When did we start trusting strangers? How the internet turned us all into influencers. Internet” *Universal McCann* <http://www.universalmccann.com/> (Consultado el 12 de febrero de 2010)
8. MERINO, José, FIERRO, Eduardo y VALENCIA, Raúl. (2010). “Candidatos opositores, los más atacados en línea. Internet”. *El Economista*. 13/06/2010. <http://eleconomista.com.mx/elecciones-2010/2010/06/13/candidatos-opositores-mas-atacados-linea> . (Consultado el 15/06/2010) Informe completo en <http://lalongheria.com/2010/06/14/campanas-internet-mexico/>
9. MAGALHÃES, Pedro (2011). “The polls in Spain and the 2011 election”. Dictada en el marco del programa “Leadership, Communication and U.S. Politics” organizado por el Departamento de Comunicación Política e Institucional del Instituto Universitario de Investigación Ortega y Gasset con la participación de docentes de la Universidad de Georgetown en octubre de 2011
10. OCIO NETWORK (2009). “Estudio de los hábitos de Internet 2009” <http://www.alejandrosuarez.es/2009/12/estudio-de-habitos-de-internet-2009-completo/> (Consultado el 15 de abril de 2012)
11. SEMIOCAST (2012). “Brazil becomes 2nd country on Twitter, Japan 3rd. Netherlands most active country.” 31/1/2012. http://semiocast.com/publications/2012_01_31_Brazil_becomes_2nd_country_on_Twitter_superseds_Japan (Consultado el 15 de abril de 2012)
12. SMITH, Aaron y RAINIE, Lee (2008). ”The internet and the 2008 election”. *Pew Internet & American Life Project*. 15/06/2008. <http://www.pewinternet.org/Reports/2008/The-Internet-and-the-2008-Election.aspx> (Consultado el 20 de noviembre de 2009)
13. SMITH, Aaron (2011). “13% of online adults use Twitter. Half of Twitter users access the service “on the go” via mobile phone”. Publicado el 01/06/2011. *Pew Internet & American Life Project. Internet*. <http://pewinternet.org/Reports/2011/Twitter-Update-2011/Main-Report.aspx> (Consultado el 20 de agosto de 2011)

14. THE COCKTAIL ANALYSIS (2011). “Observatorio de redes sociales, 3ª oleada”. 21/02/2011 http://www.slideshare.net/TCAnalysis/tca-observatorio-redes-sociales2011publico?from=ss_embed (Consultado el 8 de abril de 2012)
15. THE COCKTAIL ANALYSIS (2012). “Observatorio de redes sociales, IV oleada”. 09/04/2012. <http://www.slideshare.net/TCAnalysis/4-oleada-observatorio-de-redes-sociales> (Consultado el 20 de abril de 2012)
16. THE CONCKTAL ANALYSIS (2012). “4ª Oleada Observatorio de Redes Sociales: Las marcas empiezan a encontrar Límites.” 10/04/2012. <http://www.tcanalysis.com/2012/04/10/las-marcas-empiezan-a-encontrar-limites-en-la-utilizacion-de-las-redes-sociales/> (Consultado el 8 de abril de 2012)
17. UNIVERSAL MCCANN (2008). “When did we start trusting strangers? How the internet turned us all into influencers”. <http://www.universalmccann.com/global/knowledge/view?id=34> . (Consultado el 13 de julio de 2009)
18. ZARRELLA, Dan (2009). “ReTweets Change Everything”. Página Web de DanZarrella. *The Social media Sceintist*. (n.d.) <http://danzarrella.com/retweets-change-everything.html> (Consultado el 10 de enero de 2011)

Índices de ilustraciones y tablas

8. Índice de ilustraciones y tablas

8.1. Índice de ilustraciones

1. Ilustración 1. Visualización de la hipótesis de la influencia de Watts y Doods (2007).....	58
2. Ilustración 2. Visualizaciones de diferentes redes realizadas por Christakis y Fowler, (2010)	65
3. Ilustración 3. Ejemplos de las movilizaciones blog para “Pope case” y “Danish cartoons” de Colbaugh y Glass (2009)	66
4. Ilustración 4. Imagen de una red de usuarios de un particular en Facebook, realizada mediante la herramienta Touchgraph	70
5. Ilustración 5. Difusión de gustos musicales en una red de amigos en Facebook, de Christakis y Fowler, (2009).....	73
6. Ilustración 6. Difusión de gustos cinematográficos en una red de amigos en Facebook, de Christakis y Fowler, (2009).....	73
7. Ilustración 7. Difusión de gustos literarios en una red de amigos en Facebook, de Christakis y Fowler, (2009)	74
8. Ilustración 8. Red de conversaciones en Twitter de Gonçalves et al. (2011) 84	
9. Ilustración 9. Visualización polarizaciones subred “retweet” y subred “mention” de Conover et al. (2011).....	110
10. Ilustración 10. Visualización de la polarización de la blogosfera americana en las elecciones de 2004, de Adamic y Glance (2005)	112

11. Ilustración 11. Ejemplo del análisis Facebook de la empresa Social Backers en el seguimiento de las elecciones 2012 en EEUU.....	148
12. Ilustración 12. Ejemplo de la gráfica interactiva de la visualización de los resultados del uso de Google Trends como herramienta de predicción electoral, de Lui, Metaxas y Mustafaraj (2011).....	150
13. Ilustración 13. Figura 1 de resultados de búsquedas en Google Trends (para grupos T1 y T2) de Lui, Metaxas y Mustafaraj (2011)	151
14. Ilustración 14. Figura 2 de resultados de búsquedas en Google Trends (para grupos T1 y T2) de Lui, Metaxas y Mustafaraj (2011)	152
15. Ilustración 15. Visualizaciones de la evolución temporal de la campaña electoral de las Elecciones Autonómicas de la Comunidad de Madrid 2011 realizada por Congosto (Fernández y Congosto, 2012)	168
16. Ilustración 16. The 2010 elections: Twitter isn't a very reliable prediction tool	171
17. Ilustración 17. Facebook, Twitter Election Results Prove Remarkably Accurate.....	172
18. Ilustración 18. Captura de la página Web de Intrade en la que se presenta la predicción electoral en EEUU 2012 en una fecha determinada	174
19. Ilustración 19. Captura de la página Web de Intrade en la que se presenta la predicción electoral en México 2012 en una fecha determinada.....	175
20. Ilustración 20. Mapa de la predicción realizada por Nate Silver en las elecciones presidenciales de EEUU en 2008.....	176
21. Ilustración 21. Estimación de la victoria de las elecciones francesas 2012, utilizando mercados de predicción/apuestas, de Rothschild (2012).....	177

22. Ilustración 22. Desviaciones encuestas electorales españolas en 2008 según análisis de Pedro Magalhães (2011)	191
23. Ilustración 23. Gráficos de resultados del análisis sobre desviación de encuestas españolas realizado por GAD3	193
24. Ilustración 24. Penetración de Redes Sociales por comunidad autónoma (España), realizado por Ocio Network (2009).....	199
25. Ilustración 25. Gráficos de uso de Redes Sociales en España realizado por Fundación Orange (2011)	201
26. Ilustración 26. Gráfico de usuarios de Facebook en España (2011), por edad y sexo, realizado por Cuentamelared (2011)	203
27. Ilustración 27. Gráficos de usuarios de Facebook en España, por edad y comunidad autónoma, realizado por Zoom Labs (2012).....	204
28. Ilustración 28. Gráfico de usuarios de Facebook en España, por sexo, realizado por Michele Caivano (2012)	205
29. Ilustración 29. Gráfico de usuarios de Twitter en España, por comunidad autónoma, realizado por Adigital (2010).....	207

8.2. Índice de tablas

30. Tabla 1. Tablas de relación usuarios-influencia extraídas del trabajo de Romero et al. (2010).....	92
31. Tabla 2. Listado de usuarios con contenido más retuiteado extraída del trabajo de Asur et al. (2011)	94
32. Tabla 3. Porcentaje de audiencia de medios de comunicación en Twitter, extraída del trabajo de Golbeck y Hansen (2010)	108
33. Tabla 4. Tablas de variables de análisis empleadas para análisis de campañas en Facebook, extraídas del trabajo de Willian y Gulati (2008).....	143
34. Tabla 5. Tabla de variables de análisis empleadas para análisis de campañas en Facebook, extraídas del trabajo de Willian y Gulati (2009).....	145
35. Tabla 6. Comparación del número de seguidores en Facebook según actividad de candidatos, extraída del trabajo de de Willian y Gulati (2009).....	146
36. Tabla 7. Tablas de resultados del análisis de resultados elecciones alemanas de 2009 según Twitter, extraída del trabajo de Tumasjan et al., (2010)	155
37. Tabla 8. Tabla de clasificación de sentimientos reflejados en Twitter acerca en los discursos del Senado de EEUU, extraída del trabajo de Yu et al., (2008)	162
38. Tabla 9. Tablas de análisis sentimental y volumen de datos en Twitter durante las Elecciones Presidenciales de EEUU en 2008, extraídas del estudio de Gayo-Avello (2010).....	166
39. Tabla 10. Relación entre condiciones particulares y recursos empleados por candidatos en sus páginas Web's y su peso en relación al resultado electoral	

en las elecciones australianas de 2004, extraída del trabajo de Gibson y McCallister (2005).....	183
40. Tabla 11. Factores influyentes en el margen de error de las encuestas	190
41. Tabla 12. Relación de Variables Independientes.....	223
42. Tabla 13. Valor de Ponderación asignada a cada variable	226
43. Tabla 14. Comparación del marco muestral en las Elecciones Generales 2011	238
44. Tabla 15. Comparación del marco muestral en las Elecciones Autonómicas a la Comunidad de Madrid 2011	239
45. Tabla 16. Comparación del marco muestral en las Elecciones Autonómicas a la Región de Murcia 2011	239
46. Tabla 17. Comparación del marco muestral en las Elecciones al Parlamento de Cataluña 2010	239
47. Tabla 18. Relación de términos monitorizados en las Elecciones Generales 2011	246
48. Tabla 19. Datos globales Twitter en las Elecciones Generales 2011	247
49. Tabla 20. Resultado Twitter en las Elecciones Generales 2011	248
50. Tabla 21. Datos globales Twitter en las Elecciones Generales 2011 – Sólo menciones	248
51. Tabla 22. Resultado Twitter en las Elecciones Generales 2011 – Sólo menciones	249

52. Tabla 23. Resultado Twitter en las Elecciones Generales 2011 – Candidato	249
53. Tabla 24. Resultado Twitter en las Elecciones Generales 2011 – Sólo menciones - Candidato	250
54. Tabla 25. Resultado Twitter en las Elecciones Generales 2011 – Partido ...	250
55. Tabla 26. Resultado Twitter en las Elecciones Generales 2011 – Sólo menciones - Partido	251
56. Tabla 27. Pronósticos encuestas versus resultados Twitter en las Elecciones Generales 2011	253
57. Tabla 28. MAE Pronósticos encuestas versus resultados Twitter en las Elecciones Generales 2011	254
58. Tabla 29. Relación de términos monitorizados en las Elecciones Autonómicas a la Comundiad de Madrid 2011	255
59. Tabla 30. Datos globales Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011	255
60. Tabla 31. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011	256
61. Tabla 32. Datos globales Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011 – Sólo menciones	257
62. Tabla 33. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011-Sólo menciones.....	257
63. Tabla 34. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid - Candidato	257

64. Tabla 35. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011 – Sólo menciones - Candidato.....	258
65. Tabla 36. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011- Partido	258
66. Tabla 37. Resultado Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011 – Sólo menciones - Partido.....	259
67. Tabla 38. Pronósticos encuestas versus resultados Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011	262
68. Tabla 39. MAE pronósticos encuestas versus resultados Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011	262
69. Tabla 40. Relación de términos monitorizados en las Elecciones Autonómicas a la Región de Murcia 2011	263
70. Tabla 41. Datos globales Twitter en las Elecciones Autonómicas a la Región de Murcia 2011	264
71. Tabla 42. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011	265
72. Tabla 43. Datos globales Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Sólo menciones.....	266
73. Tabla 44. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Sólo menciones	266
74. Tabla 45. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Candidato	266

75. Tabla 46. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Candidatos – Sólo menciones	267
76. Tabla 47. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Partidos	267
77. Tabla 48. Resultado Twitter en las Elecciones Autonómicas a la Región de Murcia 2011 – Partidos – Sólo menciones	268
78. Tabla 49. Pronósticos encuestas versus resultados Twitter en las Elecciones Autonómicas a la Región de Murcia 2011	271
79. Tabla 50. MAE pronósticos encuestas versus resultados Twitter en las Elecciones Autonómicas a la Región de Murcia 2011	271
80. Tabla 51. Relación de términos monitorizados en las Elecciones al Parlamento de Cataluña 2010	273
81. Tabla 52. Datos globales Twitter en las Elecciones al Parlamento de Cataluña 2010	274
82. Tabla 53. Resultado Twitter en las Elecciones al Parlamento de Cataluña 2010	274
83. Tabla 54. Resultado Twitter en las Elecciones al Parlamento de Cataluña 2010 – Sólo menciones.....	275
84. Tabla 55. Pronósticos encuestas versus resultados Twitter en las Elecciones al Parlamento de Cataluña 2010	278
85. Tabla 56. MAE pronósticos encuestas versus resultados Twitter en las Elecciones al Parlamento de Cataluña 2010.....	278

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS DE LA INFORMACIÓN



**PREDICCIÓN ELECTORAL MEDIANTE ANÁLISIS
DE REDES SOCIALES**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR**

Montserrat Fernández Crespo

Bajo la dirección del doctor

Antonio López Vega

MADRID, 2013

86. Tabla 57. Datos globales Twitter en las Elecciones Generales 2011 – 7 días	280
87. Tabla 58. Resultados Twitter en las Elecciones Generales 2011 – 7 días ...	280
88. Tabla 59. Datos globales Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011 – 7 días.....	281
89. Tabla 60. Resultados Twitter en las Elecciones Autonómicas a la Comunidad de Madrid 2011 – 7 días	281
90. Tabla 61. Volumen de tweets en los procesos electorales monitorizados ...	282
91. Tabla 62. Porcentaje de población con derecho a voto usuaria de la red Twitter.....	283
92. Tabla 63. Relación del MAE obtenido en Twitter con el volumen de tweets por población y usuarios	286
93. Tabla 64. Relación del MAE obtenido en Twitter con el número de usuarios únicos por población y volumen de tweets.....	286